Gulchohra Mammadova
Boris Mordukhovich
Telman Aliev
Kamil Aida-zade (Eds.)

# Book of Selected Papers

3rd International Conference on

## Information Technologies and Their Applications.

ITTA 2026

April 22-24, 2026
Baku, Azerbaijan

Gulchohra Mammadova
Boris Mordukhovich
Telman Aliev
Kamil Aida-zade (Eds.)

# Book of Selected Papers

### 3rd International Conference on

## Information Technologies and Their Applications

### ITTA 2026

August 22-24, 2026
Baku, Azerbaijan

## *Editors*

Gulchohra Mammadova
Azerbaijan University of Architecture and Construction
Baku, Azerbaijan

Telman Aliev
Advisor to the President of the Azerbaijan National Academy of Sciences
Baku, Azerbaijan

Boris Mordukhovich
Wayne State University
Wayne, USA

Kamil Aida-zade
Institute of Mathematics
Baku, Azerbaijan

## *Technical editors*

Ahmadli, Nigar
Alizada, Tahir
Aydazada, Ilham
Azimov, Rustam
Gadimova, Leyla
Rzayeva, Sona

ISBN: 978-9952-8656-0-8

# Organization

**General Chair**

| | |
|---|---|
| Gulchohra Mammadova | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Program Committee Chair**

| | |
|---|---|
| Boris Mordukhovich | Wayne State University (USA) |

**Co-Chairs**

| | |
|---|---|
| Rasim Alguliyev | Institute of Information Technology (Azerbaijan) |
| Nigar Aslanova | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Honorary Chair**

| | |
|---|---|
| Telman Aliev | Advisor to the President of the Azerbaijan National Academy of Sciences (Azerbaijan) |

**Executive Secretary**

| | |
|---|---|
| Kamil Aida-zade | Institute of Mathematics (Azerbaijan) |

**Academic Secretary**

| | |
|---|---|
| Naila Musaeva | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Program Committee Members**

| | |
|---|---|
| Abdullayev, Vagif | Azerbaijan State Oil and Industry University (Azerbaijan) |
| Abdullayev, Vugar | Azerbaijan State Oil and Industry University (Azerbaijan) |
| Ahmedzade, Perviz | Ege University (Türkiye) |
| Akleylek, Sedat | University of Tartu (Estonia) |
| Aliguliyev, Ramiz | Institute of Information Technology (Azerbaijan) |
| Aliev, Fikret | Baku State University (Azerbaijan) |
| Aytaç, Aysun | Ege University (Türkiye) |
| Bagirov, Adil | Federation University (Australia) |
| Bashirov, Agamirza | Eastern Mediterranean University (Northern Cyprus) |
| Bomba, Andrii | National University of Water Management (Ukraine) |
| Boutaba, Raouf | University of Waterloo (Canada) |
| Boyd, Stephen | Stanford University (USA) |
| Cengiz, Korhan | Prince Mohammad Bin Fahd University (Saudi Arabia) |
| Chikriy, Arkadiy | V.M. Glushkov Institute of Cybernetics (Ukraine) |
| Cristofol, Michel | Aix-Marseille University (France) |
| Dursun, Mahir | Gazi Üniversitesi (Türkiye) |
| Dwivedi, Ashutosh Dhar | Aalborg University (Denmark) |
| Eppelbaum, Lev | Tel Aviv University (Israel) |
| Faraslis, Ioannis | University of Thessaly (Greece) |
| Fiaidhi, Jinan | Lakehead University (Canada) |
| Garousi, Vahid | Queen's University Belfast (United Kingdom) |

| | |
|---|---|
| Gasimov, Yusif | Azerbaijan University (Azerbaijan) |
| Górecki, Jaroslaw | Bydgoszcz University (Poland) |
| Hajiyev, Chingiz | Istanbul Technical University (Türkiye) |
| Hamzi, Boumediene | Alan Turing Institute (United Kingdom) |
| Hao, Dinh Nho | Hanoi Institute of Mathematics (Vietnam) |
| Ivković, Nikola | University of Zagreb (Croatia) |
| Jaćimović, Milojica | University of Montenegro (Montenegro) |
| Jin, Jionghua | University of Michigan (USA) |
| Kalimoldayev, Maksat | Institute of ICT (Kazakhstan) |
| Kasımzade, Azer | Ondokuz Mayıs University (Türkiye) |
| Kasimbeyli, Refail | Eskisehir Technical University (Türkiye) |
| Lee, Wai Kong | Universiti Tunku Abdul Rahman (Malaysia) |
| Lesnic, Daniel | University of Leeds (UK) |
| Litman, Amelie | Aix-Marseille University (France) |
| Lozovanu, Dmitrii | Moldova State University (Moldova) |
| Maennel, Olaf | University of Adelaide (Australia) |
| Mahjoub, Ali Ridha | Paris Dauphine University (France) |
| Mammadov, Musa | Deakin University (Australia) |
| Mamyrbayev, Orken | Institute of ICT (Kazakhstan) |
| Mardani, Ali | Bursa Uludag University (Türkiye) |
| Melikov, Agassi | Baku Engineering University (Azerbaijan) |
| Mihajloska, Hristina | Ss. Cyril and Methodius University in Skopje (Macedonia) |
| Molina-Moreno, Valentín | University of Granada (Spain) |
| Nabiyev, Vasif | Karadeniz Technical University (Türkiye) |
| Nasiboglu, Efendi | Dokuz Eylül University (Türkiye) |
| Nesterov, Yurii | Université Catholique de Louvain (Belgium) |
| Normatov, Ibrokhimali | National University of Uzbekistan (Uzbekistan) |
| Nunez-Cacho, Pedro | University of Jaén (Spain) |
| Nuriyev, Urfat | Karabakh University (Azerbaijan) |
| Oproescu, Mihai | University of Pitesti (Romania) |
| Parasidis, Ioannis | University of Thessaly (Greece) |
| Pogoriliy, Sergey | Taras Shevchenko National University (Ukraine) |
| Polyak, Roman | George Mason University (USA) |
| Providas, Efthimios | University of Thessaly (Greece) |
| Rahimov, Rustam | Tashkent State Transport University (Uzbekistan) |
| Ramadan, Rabie A. | University of Nizwa (Oman) |
| Rolbiecki, Roman | Bydgoszcz University (Poland) |
| Rustamov, Samir | ADA University (Azerbaijan) |
| Sabah, Mohammed | Lakehead University (Canada) |
| Semenov, Volodymyr | Taras Shevchenko National University (Ukraine) |
| Sidorenko, Sergii | National Technical University of Ukraine (Ukraine) |
| Sladkowski, Aleksander | Silesian University of Technology (Poland) |
| Slodicka, Marian | Ghent University (Belgium) |
| Stetsyuk, Petro | V.M. Glushkov Institute of Cybernetics (Ukraine) |
| Tadumadze, Tamaz | Tbilisi State University (Georgia) |
| Van Bockstal, Karel | Ghent University (Belgium) |

**Organizing Committee Chair**

| | |
|---|---|
| Israfil Mamedov | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Co-Chairs:**

| | |
|---|---|
| Elkhan Sabziev | Institute of Mathematics (Azerbaijan) |
| Sabir Mamedov | Azerbaijan University of Architecture and Construction (Azerbaijan) |
| Sarkhan Talibzada | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Organizing Committee Members**

| | |
|---|---|
| Abdullayev, Nijat | Azerbaijan University of Architecture and Construction |
| Abdullayeva, Gulchin | Institute of Mathematics (Azerbaijan) |
| Ahmadli, Nigar | Institute of Mathematics (Azerbaijan) |
| Aliyev, Alakbar | Baku State University (Azerbaijan) |
| Aliyeva, Aygun | Institute of Mathematics (Azerbaijan) |
| Alizada, Tahir | Institute of Mathematics (Azerbaijan) |
| Apollo, Magdalena | Gdańsk University of Technology (Poland) |
| Ashrafova, Yegana | Baku State University (Azerbaijan) |
| Azimov, Rustam | Institute of Mathematics (Azerbaijan) |
| Bashirov, Rza | Eastern Mediterranean University (Northern Cyprus) |
| Chaudhary, Prince Shiva | IEEE Senior Member (USA) |
| Gelashvili, Otar | Technical University (Georgia) |
| Gasimov, Vagif | Baku Engineering University (Azerbaijan) |
| Guliyev, Samir | Azerbaijan State Oil and Industry University (Azerbaijan) |
| Guseynov, Sharif | Riga Technical University (Latvia) |
| Gürsoy, Arif | Ege University (Türkiye) |
| Huseynova, Rana | Azerbaijan University of Architecture and Construction |
| Isayev, Mazahir | Azerbaijan University of Architecture and Construction |
| Ibrahimov, Bayram | Azerbaijan Technical University (Azerbaijan) |
| Kazim-zade, Aydin | Azerbaijan University of Architecture and Construction |
| Khurshudov, Dursun | Azerbaijan University of Architecture and Construction |
| Kuspangaliyev, Bolat | Institute of Architecture and Civil Engineering (Kazakhstan) |
| Leonov, Sergey | National Technical University (Ukraine) |
| Mammadov, Vusal | Institute of Mathematics (Azerbaijan) |
| Mammadova, Masuma | Institute of Information Technology (Azerbaijan) |
| Nuri, Fidan | Yaşar University (Türkiye) |
| Orujov, Orkhan | Azerbaijan University of Architecture and Construction |
| Ordin, Burak | Ege University (Türkiye) |
| Pashayev, Adalat | Institute of Mathematics (Azerbaijan) |
| Rahimov, Anar | Institute of Mathematics (Azerbaijan) |
| Rupp, Florian | Kutaisi International University (Georgia) |
| Rzayev, Ramin | Institute of Mathematics (Azerbaijan) |
| Rzayeva, Narmin | Azerbaijan Technical University |
| Rzayeva, Sona | Institute of Mathematics (Azerbaijan) |
| Sadigov, Aminaga | Institute of Mathematics (Azerbaijan) |

| | |
|---|---|
| Sattarova, Ulkar | Ministry of Science and Education (Azerbaijan) |
| Serovajsky, Simon | Al Farabi Kazakh National University (Kazakhstan) |
| Stovba, Viktor | V.M.Glushkov Institute of Cybernetics (Ukraine) |

**Technical support group**

**Co-Chairs:**

| | |
|---|---|
| Reyhan Akbarli | Azerbaijan University of Architecture and Construction (Azerbaijan) |
| Reshad Ismibayli | Azerbaijan University of Architecture and Construction (Azerbaijan) |

**Members**

| | |
|---|---|
| Ahmadli, Nigar | Technical Editor (Institute of Mathematics, Azerbaijan) |
| Aliyeva, Aygun | Treasurer of the Conference (Institute of Mathematics, Azerbaijan) |
| Alizada, Tahir | Contact Person for Springer (Institute of Mathematics, Azerbaijan) |
| Aydazada, Ilham | Technical Editor (Institute of Mathematics, Azerbaijan) |
| Mammadov, Vusal | Webmaster (Institute of Mathematics, Azerbaijan) |

# Table of Contents

**Decision-Making and Optimal Control** 1–2

*Esmanur Alican, Caner Özcan, İsmail Rakıp Karaş*
A Review of Graph-Enhanced Augmented and Mixed Reality 3–13

*Okkes Tolga Altinoz*
Redesigning the Particle Swarm Optimization Algorithm as a Reinforcement Learning Methodology 14–27

*Kira Dmitrieva, Majid Abbasov*
Application of the RRT Algorithm to Solve Control Theory Problems 28–34

*Samia Akter Erin, Firuja Tasneem, Mim Mony, Mst. Umma Nourin Sawon, Mohammad Nyme Uddin*
$CO_2$ Concentration Prediction in a Multi-Functional Semi-Open University Auditorium Using Regression-Based Machine Learning Models 35–50

*Melih Eşiyok, Okan Uluturk, Berke Bekar, Israfil Isakhli, Efendi Nasiboglu*
Detection of the Braess Paradox in Signalized Road Networks via Traffic Flow Simulation 51–63

*Khanlar Gamzaev, Sevinc Kerimova*
Computer Simulation of the Elastic Regime of the Development of a Single-Well Reservoir with an Unknown Initial State 64–72

*Abdulvahit Karail, Yasin Ortakcı*
Impact of Optimizer: Comparative Analysis of Loss Functions in Different Models 73–88

*Nermin Kartli*
A Dynamic Programming Approach for the Fuzzy Minimum Cost Maximum Flow Problem in the Multi-Stage Networks 89–102

*Makhamadaziz Rasulmukhamedov, Gulmira Mirzaeva, Nomaz Mirzaev, Nuraddin Gafforov*
Extraction of Features Characterizing Keystroke Dynamics in the User Identification problem 103–112

*Heydar Suleyman Rzayev, Arzu Agababa Akhundov, Natavan Mammad Khasayeva, Mehriban Bakhtiyar Mammadova*
Intelligent Information-Measuring System for Scanning Liquid Flows 113–124

**ICT and Intelligent Systems in Sociology and Economics** 125–126

*Svetlana Allahverdiyeva, Vafa Jafarova*
Noise in Big Cities and Methods of Combating It 127–138

*Furkan Atban, Muhammed Yusuf, Cüneyt Bayılmış*
Quantum Transfer Learning with Pretrained CNN Features for Multi-Class Medical Image Classification 139–152

*Shavkat Fazilov, Shukhrat Mamadjanov*
System Analysis of Acne Diagnostics Using Artificial Intelligence 153–168

*Shahla Gahramanova, Kamal Ibrahimov*
From Caspian Shores to Mountain Peaks: ICT-Powered Tourism Resilience in Azerbaijan 169–183

*Kalen Meiirzhan Kanatuly*
Digital Technologies and Their Impact on the Optimization of Business Processes in the Company: Focus on RPA 184–194

*Tam Tran Luu, Luan Trong Nguyen*
Enhancing Learning Effectiveness and Satisfaction through Game-Based Learning in Higher Education 195–208

*Afat Mammadova Mammadhuseyn*
Digital Infrastructure and Regional Economic Development: Comprehensive Assessment of Digitalization in the Regions of Azerbaijan 209–220

*Isa Muradov, Natig Mahamaliyev*
Digital Sovereignty in Tourism Marketing: Data Governance and Cybersecurity Challenges in Emerging Markets 221–232

*Nooshin Nemati, Emrah Hancer, Akın Atakan Goren, Refik Samet*
A Comparative Study of YOLO-based Architectures for Mitosis Detection in Histopathology Images 233–244

*Rafail Rzayev, Elman Ibishov, Aliagha Gasimov, Arzu Safarova*
Artificial Intelligence and Higher Education Quality: Survey Evidence from UNEC 245–255

**ICT in Pattern Recognition** 256–257

*Maha Albayati, Oğuz Fındık*
Ensemble Strategies for Turkish Legal Extractive Summarization Using Semantic Voting and ROUGE-Weighted Averaging 258–271

*Yusuf Evren Aykaç, Merve Özkan, Refik Samet*
Lexicon-Augmented Explainable Aspect-based Sentiment Analysis for Azerbaijani Finance and Business Text 272–285

*Kübra Kovayçin, Ece İrem Şişer, Derda Sina Günay, Yusuf Evren Aykaç, Refik Samet*
Section-Aware Retrieval for Turkish Case Law: A Case Study on Scanned Court Decisions 286–298

*Anvarkhon Majidov*
A Fuzzy Neural Network Approach for Intelligent Environmental Monitoring and Anomaly Detection in IoT Networks 299–312

*Nomaz Mirzaev, Johongir Urinboev, Sayyora Ibragimova, Gulmira Mrzaeva, Azizbek Tillavoldiev*
The Problem of Forming a Feature Space for Speaker Recognition 313–330

*Süleyman Nurullah Adahi Şahin, Abdulkadir Özden, Hakan Kocaman, Abdurrahman Korkmaz, Yavuz Halimergün, Cem Özkurt*
Mapping Mobility Justice: A Deep Learning and GIS-Based Framework for Micro-Scale Sustainable Mobility Policy Assessment 331–340

*Buse Sarıçayır, Caner Özcan, Yasin Ortakçı*
A Comparative Study of Word Embedding Techniques for Turkish Sentiment Analysis 341–357

**Industry and Technical Tools** 358–359

*Konul Gafarbayli, Bayram Aslanov*
Smart Buildings, Smart Villages, and Smart Cities: Applications of Digital Technologies in Architecture 360–372

*Azer Kasimzade, Varol Koç, Furkan Gunday, Hasan Yılmaz, Emin Nematli*
A New Approach in Structural Control: An Industrial Application for the Tallest Building in the Caucasus 373–395

*Hoang Duc Long*
Synthesis of Terminal Synergetic Controller Based on Nonlinear State Observer for 2-DOF Helicopter System 396–412

*Nizami Nagiyev, Sabina Aliyeva, Elbay Aliyev*
Digital Transformation of Rural Settlement Systems in Azerbaijan 413–422

*Bensu Melis Orhan, İrem Kayaarslan*
A Search and Rescue Robot for Under Rubble Victim Detection Using Thermal and Sensor Based Systems 423–437

*Yaşar Nuri Sevgen*
An Energy-Based Adaptive Modulation Framework for Bandwidth-Efficient Information Transmission 438–451

*Ilhami Suleymanov, Nazli Muradova*
An Intelligent Dust Detection and Protection System (DDPS) for Autonomous Management of Photovoltaic Performance 452–466

*Enrico Zacchei, Md. Munir Hayet Khan*
Advanced Modelling of Chloride Penetration in Concrete Considering Time- and Space-Dependent Diffusivity 467–478

*Enrico Zacchei, Zhiyuan Wang*
Clustering-Based Post-Processing of Experimental Data for Double Glass Units 479–490

*Nurtakin Guliyeva Shukurova*
Simulation-based Comparative Study of PROFIBUS and PROFINET Topologies 491–502

# Decision-Making and Optimal Control section

# A Review of Graph-Enhanced Augmented and Mixed Reality

Esmanur Alican[0009-0003-0575-7766], Caner Özcan [0000-0002-2854-4005], and İsmail Rakıp Karaş [0000-0001-5934-3161]

[1] Karabuk University, Karabuk 78000, Türkiye
esmnralicann@gmail.com

**Abstract.** Augmented Reality (AR) and Mixed Reality (MR) are the leading technologies that provide us with new ways to interface between the digital world and real spaces. This paper presents a literature review on the evolving intersection between graph theory and AR/MR, and how graph-based approaches advance these technologies. We present an analysis of the literature on this topic published between the years 2019 and 2024. This time frame captures articles discussing a variety of applications including, but not limited to, cultural heritage tourism, navigation, gaming, education, and collaborative work. The review also describes the functionalities of graphs within the AR/MR environments, including visualization, spatial understanding, scene rendering, network cutting, and user engagement. We assess the opportunities and challenges of using graph-based approaches, summarize some of the research areas currently on the agenda, and suggest some possibilities for future work in this area.

**Keywords:** Augmented Reality, Mixed Reality, Graph Theory

## 1 Introduction

The development of new technologies has led to the design of applications to solve the problems of everyday life. In fact, when designing such applications, it is crucial to design using suitable visualization criteria. People's interaction with the world around them, as well as their ability to understand and process information about this world, is evolving significantly. The integration of augmented reality (AR) and mixed reality (MR) and graph technology could change how we engage with the world around us. In many fields, there is growing interest in exploring approaches that utilize mixed reality (MR) and augmented reality (AR) technologies [1]. Mixed and augmented reality have become embedded in our daily activities. Major developments in areas such as navigation [2], social networking [3], entertainment [4], remote sensing, and engineering [5] demonstrate the promising potential of AR and MR technologies.

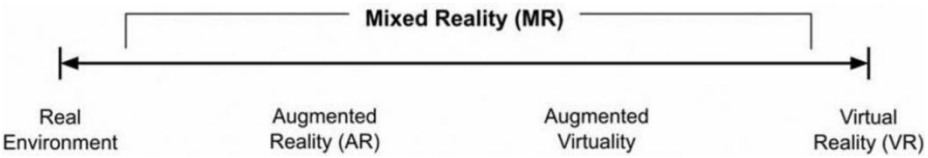



**Fig. 1.** The Reality-Virtuality Continuum, as proposed by Milgram

A widely accepted definition of AR is provided by Azuma, who defines it as a technology that integrates real and virtual objects that interact in real time and are embedded in the real world [7]. AR implies boosting, i.e. the putting into use of visible spectrum images, video, and 3D models into real-world environments and surroundings of the environment perceived by the user. This technology helps to erase the boundaries of the virtual and the real by enabling people to use digital content in the actual world [8].

MR, on the other hand, takes it a step further and combines digital information with the physical environment, allowing for more interactive experiences [9]. MR is able to merge the classical notion about physical and virtual worlds being two opposites into a single strong sense [10]. Together with other sensing capabilities, objects may be placed anywhere along the reality-virtuality continuum to not differentiate them clearly. This fusion, based on a single design of a real object and a virtual object, seamlessly integrates the real and virtual perceptions in MR. For this reason, the user's understanding of the MR reality is layered and complex since the divisions between physical things and pictorial representations are blurred.

Euler's work on the Seven Bridges of Königsberg is considered the beginning of graphs [11]. He stated that graphs are mathematical structures that represent relationships between entities by using nodes (or vertices) to represent entities and edges to represent the connections between them. Recent work has improved Euler's work, showing how graphs can be leveraged in augmented and mixed reality systems to enhance real-time data visualization and improve user interactions in immersive environments [12]. The role of graphs in augmented and mixed reality systems is to organize and contextualize digital components by representing relationships and structures visually and interactively.

This paper is structured as follows: In Section 1, an overview of the research problem and aims of this study is given and describes the framework for the analysis carried out thereafter. In Section 2, the works relevant to the topic are discussed. Section 3 describes the research methodology including the techniques, and the conclusion is presented in Section 5.

## 2 Related Works

After a brief introduction to AR, MR and Graphs, this paper presents an in-depth review of the existing literature and analyses related applications and papers.

### 2.1 Augmented Reality Applications with Graphs

Augmented Reality (AR) is becoming a game changing technology in which the digital information is integrated with the real world in such a way as to enhance the experiences of users. An enormous development in the capacity of AR applications to present and handle complex data was especially brought about by the combination of graph theory and graph algorithms.

Qassimi and Abdelwahed [13] focus on information overload which is an aspect of concern in Cultural Heritage Tourism, and it is made worse by the many platforms that allow users to post content. Their paper, presents an innovative method for suggesting appropriate sites of the cultural heritage to a sightseerer. Noting the drawbacks of folksonomy with uncontrolled vocabulary and the resulting ambiguity of meaning, the authors offer a solution based on semantic graph. Their system is designed to provide an advanced semantic/meta description graph for each site, and the graphs are then used to map semantic relations between tags of similar sites and between the sites themselves. As a result, recommendations are expected to give preferences in suggestions that are more relevant rather than that the artificial intelligence using folksonomy concepts. The authors indeed found some promising results from their experimental evaluations, showing the efficiency of such semantic graph-based recommender systems in the context of cultural heritage tourism.

Büschel et al. [14] investigates the effectiveness of visualizing graphical link attributes in mobile augmented reality (MAR) environments. The study focuses on how different visual coding techniques for connectivity attributes such as color, thickness, and animation affect user comprehension and task performance in MAR graph visualizations. The research will focus on the interaction between these graph visualization principles and what MAR uniquely offers, by spatial mapping and real-world context, to identify the optimal strategy for representing intricate relationships within augmented overlays. The research contributes to knowledge on how graph-based information can be conveyed effectively in MAR, leading toward the creation of more intuitive and insightful graph-based AR applications.

Li and Wang [15]'s interactive AR graph demonstrates the effectiveness of graph theory in depicting the complex interactions between Chinese painters. In this case, the 3D space can be defined as an AR space in which these networks can be visualized. Users can immerse themselves in the environment and comprehend how painters interacted and the numerous trends that shaped art history. In the same vein, Schwajda et al [16] propose an orthogonal visuals approach where 2D graph data is transformed into 3D AR space for better interactivity and visual understanding. This study aims to investigate how different types of graphical representations, including directed graphs and weighted graphs, affect the understanding of complex datasets and the various AR tools they utilize in their respective cases. They also highlight the potentially powerful capabilities of AR when it comes to accurately depicting the data behind complex datasets. In addition, they provide quantitative evidence of the effectiveness of the proposed AR visualization, considering the effects on user performance and satisfaction in terms of metrics.

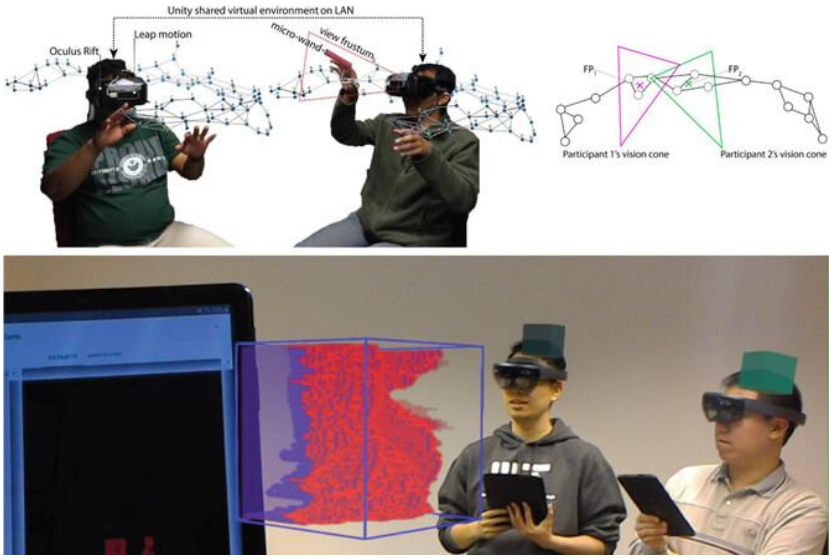



**Fig. 2.** Examples of XVA systems with cross-device and collaborative features are starting from above, Cordeil et al's analysis evaluation of immersive collaborative networks‘ evaluation (top, © 2016 IEEE. Reprinted, with permission, from Cordeil et al.[CDK*17] ); follow up figure caption This is an illustration of a volumetric data visualization and analysis prototype described by Sereno et al. (bottom, © M. Sereno, L. Besançon and T. Isenberg, reprinted, with permission, from Sereno et al.[SBI19]).

Lampropoulous et al. [9] provide a gathering where in-depth review of ar functionality improvement with the use of learnt deep, networks, semantic web and knowledge graphs. Knowledge graphs represent relationships between objects and concepts and are introduced as a contextual background to AR applications so as to make them meaningful. With a focus on different application areas such as museum guidance, education, and healthcare. The study sheds light on how these technologies can bewenchical to Augmented Reality Experiences. The merger if knowledge graphws especially enables the users to know more about real-life objects and the concepts behind them making one more immersed into the AR applications. AR applications especially multi-users and collaboration’s have great expectations for the network performance. Pereira et al [17] implemented Arena, an AR edge networking architecture open out to enhancing network performance, and enlarges networks of AR edge applications. Arena uses graph based means in network resource optimization to cut latency. The research describes and Animates the design and Aggregation of Arena with evidence through simulation of performance enhancement. Contributing to this field, Jeon and Woo [18] eCAR architecture addresses the limits of edge performance in optimizing collaborative AR experiences. eCAR uses graph-based methods to improve efficiency and scalability. eCAR addresses the problems of helping a large number of participants experience a common AR fidelity using remote forces. Since several networks carry an AR device and global edge servers that use graphs as data movement devices, a graph will likely identify connections. Alternatively, it is possi-

ble that nodes represent users and the virtual objects they interact with in the shared space of the AR that honors a graph. Since directed graphs will show how edges correspond to interaction, the paper seems to cover the main ideas and principles laid out in directed edge computing systems based on graphs.

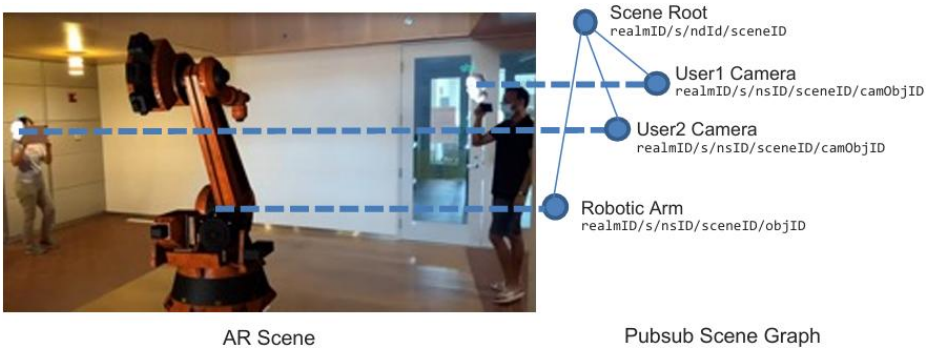



**Fig. 3.** ARENA Scene The objects in an ARENA Scene are indirectly networked via a PubSub bus where each object is controlled by a topic endpoint. The topic hierarchy is partitioned by Realm (RealmID), owner namespace (nsID), Scenes (sceneID) and objects within the Scene (camObjID or objID)[17].

Liu et al. [19] present an approach for the generation of AR gaming sceneries, which combines the ideas of spatial design and reality. Their approach implements a graph-based scene representation which assists in the development of more adaptive and interactive gaming environments, enabling the use of sophisticated graphical tools for the purpose of creating animated structures based on AR for the entertainment industry. The experimentation proved the viability of the proposed adaptive method for game development using AR, demonstrating a prototype of implementation. Qiu et al. [20] describe NavMarkAR, an augmented reality navigation system that uses landmarks to assist users from elderly populations who wish to improve their spatial cognition. In their work, they used graphs to represent the landmark relations that form the AR-enhanced mental maps, helping the them use augmented reality. One of the participants' evaluations conducted in the study showed that NavMarkAR helped elderly subjects to enhance spatial cognition and orientation abilities. Graph-based AR will evolve in the future by looking at more efficient graph algorithms and data structures and new interaction modes. Of great importance is still performance optimization, without which the use of mobile devices remains a great challenge. Graph based augmented reality applications are very mobile and so user interface and user experience design are also very important. Additionally, bearing in mind Artificial Intelligence and Deep Learning as well as Computer Vision makes it possible to make AR systems smarter and more context-aware. In conclusion, Graph-based techniques stand out in enhancing the efficiency and flexibility of the developed AR applications because the research in this area would be further developed to revolutionize the user interactions with either real or virtual worlds.

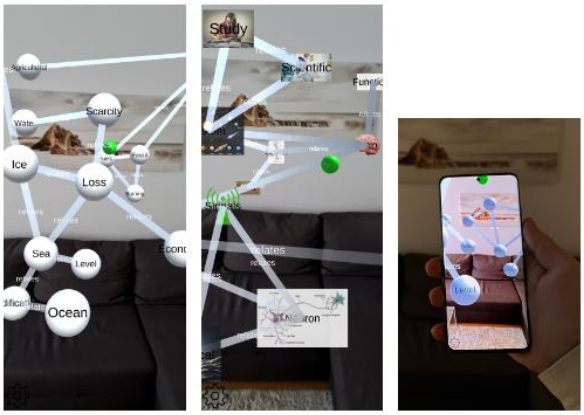
According to Belcher et al. [21], there is an improvement in the use of Augmented Reality (AR) in the visualization of complex graphs in three dimension. These systems allow users to use AR to visualize graph structures overlaid on the real environment and to interact with the graphs themselves. Such systems employ a tracked handheld display which renders the graph such that the user can view the graph and even move the graph nodes and edges in the 3D space. The authors of this paper argue that this approach overcomes some of the limitations inherent in viewing the graph in a 2D space, most notably providing spatial cues and enabling comprehending the complex relationships of the graph apace in a natural manner. The paper describes a prototype implementation demonstrating the proof of concept for the idea and usability aspects of augmented reality in graph visualization, especially for software visualization and social network analysis.

In their work, Tahara et al. [22] elaborated the concept of Retargetable AR which makes use of a 3D scene graph to enhance augmented reality addressable to indoor spaces. They put forward the idea of directed acyclic graphs (DAGs) as a means of establishing the interconnection and organizational structure between objects in a scene, thus rendering the effective and appropriate visualization and interaction of the virtual options within the context in which they are placed. The retrieval of a scene graph where locations of virtual objects are placed with consideration of their spatial semantics attributes are not random since the system provides high-level attributes related to the virtual and its orientation. This means that objects for the tangible world are virtually projected onto the real world in perfect compatibility that, for instance, virtual furniture can be projected on a real table, and virtual decorations can be put or painted on the walls. In addition, the retarget ability feature allows transferring the AR for other indoor spaces just by using another scene graph, therefore increasing the applicability and universality of the system. Their findings highlight the relative effectiveness of the Graphics representation in AR as they are able to create meaning coherent sites using the AR interface.

### 2.2. Mixed Reality Applications with Graphs

Graph-based techniques are relatively new in Mixed Reality (MR), and they are contributing to the enhancement of user experience and resolving some of the technical aspects related to MR. Several key studies highlight the diverse ways in which graph structures and algorithms can enhance MR experiences.

Hensen et al. [23] conducted a study which aimed at creating interactive visualizing learning structures with the help of knowledge graphs in MR. They made knowledge graph nodes and edges as 3D objects and relationships in the MR environment, so that users were able to touch abstract ideas. Such interaction allows users to get a better comprehension of difficult topics in comparison to ordinary visualization. This study confirmed the effectiveness of knowledge graph implementation into MR in designing the environments that deinstitutionalize learners and make them more engaged in the learning process.




(a) Sphere representation. (b) Image representation. (c) App on a smartphone.

**Fig. 4.** Two KG representations and a student using the app on a smartphone[23].

Localization is one of the most important elements when considering MR interfaces due the fact that MR has to be accessed from different devices. Addressing the issue of cross-device localisation, Osipov, Ostanin and Klimchik [24] attempted to compare different point cloud registration algorithms with some graph-based methods, such as FIGRA.Their objectives in the research were to find out reliable algorithms which could facilitate effective global localization in MR.

Hernández et al. [25] explored the development of thermodynamics-informed neural networks for realizing physical realism in MR simulations. As dealing with graph neural networks (GNNs) in their particular application raises more questions than it answers, their research demonstrates the promise of embedding physical constraints in the learning of neural networks. This method is intended to produce more reliable and realistic simulations of MR, in areas such as engineering and design, whereby realistic physical interactions need to be modeled in great detail.

Mu et al. [26] designed a spatial location and tracking model for MR using the concepts of SLAM algorithms. This model incorporates a visual slam where visual images are collected and used for creating a three-dimensional perspective of the surroundings and the position of the MR device in it. Although the specificity of the scope remains a prominent concern, particularly in assessing the accuracies and real-time aspects of these algorithms without comparing them against other SLAM algorithms, these are still promising findings. Future studies should aim to improve algorithms and test their effectiveness in many scenarios so that they can be used as scalable in higher-order MR environments.

Regardless, the paper remains a valuable document for researchers wanting to study this emerging interdisciplinary area and as noted in Table 1's analysis of various applications, aptly highlights future research opportunities.

**Table 1.** Mode Related works

| Reference | Study Topic | Technology | Chart Type |
|---|---|---|---|
| Qassimi & Abdelwahed [13] | Cultural Heritage Tourism | AR | Semantic Graph |
| Büschel [14] | Link Visualization in Mobile AR | AR | Visual Encoding |
| Li & Wang [15] | AR-Based Artist Network | AR | 3D Graph |
| Schwajda et al. [16] | 2D to 3D AR Conversion | AR | Directed Graph |
| Lampropoulos et al. [9] | Information Graphs in AR | AR | Knowledge Graph |
| Pereira et al. [17] | Arena: Edge Computing AR | AR | Graph Optimization |
| Jeon & Woo [18] | eCAR: Edge Computing Optimization | AR | Graph-Based System |
| Liu et al. [19] | AR Gaming Environment | AR | Graph-Based Scene |
| Qiu et al. [20] | NavMarkAR: AR Navigation | AR | Graph-Based Routing |
| Belcher et al. [21] | 3D AR Graph Visualization | AR | 3D Graph |
| Tahara et al. [22] | Retargetable AR | AR | DAG (Directed Acyclic Graph) |
| Hensen, Rechtmann & Neumann [23] | Information Graphs in MR | MR | Knowledge Graph |
| Osipov, Ostanin & Klimchik [24] | Positioning in MR | MR | Point Cloud |
| Hernández et al. [25] | Physical Simulation in MR | MR | GNN (Graph Neural Network) |
| Mu et al. [26] | SLAM Tracking in MR | MR | Visual SLAM |

## 3 Discussion

This literature review is helpful in understanding the confluence of graph theory and augmented/mixed reality (AR/MR) technologies focusing on research published between 2019 and 2024. The authors are successful in demonstrating the cross-domain applications of graph approaches, such as in cultural heritage, navigation, gaming, and collaborative work. The review describes the main functionalities graphs provide in AR/MR environments: Visualization and understanding of space, rendering of scenes, optimization of networks, and participation level of users. Nonetheless, some challenges regarding performance optimization for mobile devices, interface design, and the application of AI and computer vision are also pointed out. It is clear that the variety of applications is a notable strength; however, the lack of a comprehensive com-

parative analysis of different graph-based approaches and their performances across applications is a missed opportunity for critical appraisal. Despite these shortcomings, the paper works well as a starting point for researchers wanting to delve into the interdisciplinary domain while outlining the optimum research strategies, such as the formulation of efficient graph algorithms, the design of interaction models, and customizing AR/MR content.

## 4 Conclusion

This paper has demonstrated both the almost ubiquitous nature and the use of graph theoretic structures in the context of LBSN research. Although no of the papers under review set out primarily to study the topology of any graph, it is important to realize that the use of different graph algorithms and structural designs greatly enhance the quality of such research. Researchers across different fields of study often employ various forms of graph structures in solving particular problems and developing their contributions in LBSN. In order to grasp the scope of research activity in this field better, the classification of the articles based on the types of graph structures used is especially helpful. Concepts and structures of graphs and their applications in this case include heterogeneous graphs made up of nodes and edges representing various components and relationships, implicit spatial networks that depict the geographical context of location-based services, graphs representing information flow and how such information flows within a social network, and graphs evolved out of specific requirements in LBSN areas. This classification not only enables one to see how graph structures are employed and used in LBSN studies, but also emphasizes the range of approaches that scholars take. Each category corresponds to a particular view as to how graph theory could further clarification of social relatedness with regard to space – and only space – indicating the complex use of graph in this extremely active area of thinking.

## References

1. Schmalstieg, D.: Augmented Reality, Principles and Practice. Addison-Wesley Professional (2016).
2. Papadopoulos, T., Evangelidis, K., Kaskalis, T.H., Evangelidis, G., Sylaiou, S.: Interactions in augmented and mixed reality: an overview. Applied Sciences 11(18), 8752 (2021).
3. Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., Stamminger, M., Nießner, M., Theobalt, C.: State of the art on monocular 3D face reconstruction, tracking, and applications. In: Computer Graphics Forum, vol. 37, pp. 523–550. John Wiley & Sons Ltd., Hoboken (2018).
4. Cordeiro, D., Correia, N., Jesus, R.: ARZombie: A mobile augmented reality game with multimodal interaction. In: 2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN), pp. 22–31. IEEE (2015).
5. Ladwig, P., Geiger, C.: A literature review on collaboration in mixed reality. In: Smart Industry & Smart Education: Proceedings of the 15th International Conference on Remote

Engineering and Virtual Instrumentation 15, pp. 591–600. Springer International Publishing (2019).
6. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems 77(12), 1321–1329 (1994).
7. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6(4), 355–385 (1997).
8. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using augmented reality in education from 2011 to 2016. In: Innovations in Smart Learning, pp. 13–18. Springer (2017).
9. Lampropoulos, G., Keramopoulos, E., Diamantaras, K.: Enhancing the functionality of augmented reality using deep learning, semantic web and knowledge graphs: A review. Visual Informatics 4(1), 32–42 (2020).
10. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. In: Telemanipulator and Telepresence Technologies, vol. 2351, pp. 282–292. SPIE (1995).
11. Euler, L.: Solutio problematis ad geometriam situs pertinentis. Commentarii Academiae Scientiarum Petropolitanae, 128–140 (1741).
12. Ferreira, V., Portela, F., Santos, M.F.: A practical solution to synchronise structured and non-structured repositories. In: Trends and Applications in Information Systems and Technologies: Volume 4, pp. 356–365. Springer International Publishing (2021).
13. Qassimi, S., Abdelwahed, E.H.: Towards a semantic graph-based recommender system. A case study of cultural heritage. JUCS: Journal of Universal Computer Science 26(7) (2021).
14. Büschel, W., Vogt, S., Dachselt, R.: Investigating link attributes of graph visualizations in mobile augmented reality. In: Proceedings of the CHI 2018 Workshop on Data Visualization on Mobile Devices, vol. 18. MobileVis (2018).
15. Li, J., Wang, Z.: An interactive augmented reality graph visualization for Chinese painters. Electronics 11(15), 2367 (2022).
16. Fröhler, B., Anthes, C., Pointecker, F., Friedl, J., Schwajda, D., Riegler, A., Heinzl, C.: A survey on cross-virtuality analytics. In: Computer Graphics Forum, vol. 41, no. 1, pp. 465–494 (2022).
17. Pereira, N., Rowe, A., Farb, M.W., Liang, I., Lu, E., Riebling, E.: Arena: The augmented reality edge networking architecture. In: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 479–488. IEEE (2021).
18. Jeon, J., Woo, W.: eCAR: Edge-assisted collaborative augmented reality framework. arXiv preprint arXiv:2405.06872 (2024).
19. Liu, J., Zhang, R., Sawabe, T., Fujimoto, Y., Kanbara, M., Kato, H.: Integrating spatial design with reality: An AR game scene generation approach. In: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 715–716. IEEE (2024).
20. Qiu, Z., Ashour, M., Zhou, X., Kalantari, S.: NavMarkAR: A landmark-based augmented reality (AR) wayfinding system for enhancing spatial learning of older adults. arXiv preprint arXiv:2311.12220 (2023).
21. Belcher, D., Billinghurst, M., Hayes, S.E., Stiles, R.: Using augmented reality for visualizing complex graphs in three dimensions. In: The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings, pp. 84–93. IEEE (2003).
22. Tahara, T., Seno, T., Narita, G., Ishikawa, T.: Retargetable AR: Context-aware augmented reality in indoor scenes based on 3D scene graph. In: 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 249–255. IEEE (2020).

23. Hensen, B., Rechtmann, A., Neumann, A.T.: Mixed reality learning visualizations using knowledge graphs. In: International Conference on Web-Based Learning, pp. 147–154. Springer Nature Singapore (2023).
24. Osipov, A., Ostanin, M., Klimchik, A.: Comparison of point cloud registration algorithms for mixed-reality cross-device global localization. Information 14(3), 149 (2023).
25. Hernández, Q., Badías, A., Chinesta, F., Cueto, E.: Thermodynamics-informed neural networks for physically realistic mixed reality. Computer Methods in Applied Mechanics and Engineering 407, 115912 (2023).
26. Mu, X., Lin, S., Duan, R.: Construction of spatial location and tracking model of mixed reality based on SLAM algorithm. In: 2024 International Conference on Telecommunications and Power Electronics (TELEPE), pp. 880–884. IEEE (2024).

# Redesigning the Particle Swarm Optimization Algorithm as a Reinforcement Learning Methodology

Okkes Tolga Altinoz[1][0000−0003−1236−7961]

Ankara University, Ankara 6830, Turkiye taltinoz@ankara.edu.tr

**Abstract.** Modern optimization algorithms consist of algorithms that have been developed, improved, and expanded for different problems with specific properties as extensive, stochastic, noise multi, many, bi-level, as well as traditional optimization problems. These relatively complex modern algorithms are mainly based on traditional numerical methods. Their complex structures are increasingly being extended with learning-based methods as they can be used in conjunction with machine learning methods, particularly reinforcement learning, and are also used within these methods. These algorithms can be evolutionary or nature-inspired algorithms. Among nature-inspired algorithms, one of the most widely used is the Particle Swarm Optimization (PSO) algorithm. The PSO algorithm has been used in conjunction with machine learning algorithms for solving the problem-oriented applications; and Reinforcement Learning as a machine learning method contains potential to use with PSO. Therefore, in this study, the PSO algorithm will be redesigned as a reinforcement learning framework. The PSO in a different framework will be evaluated on benchmark problems where the innovation suggestions will be made regarding the parameters and formulations of the algorithm according to their areas of use, its performance will be evaluated, and improvement suggestions will be presented.

**Keywords:** Particle Swarm Optimization · Reinforcement Learning · Learning-based Optimization · Machine Learning · Single-objective · Optimization.

## 1 Introduction

Teaching–learning-based optimization methods are known for simplicity and fast convergence with exploration-exploitation balance and preferred engineering applications. Optimization algorithms have been evaluating with the machine learning algorithm to form learning-based optimization algorithms. Since learning is essential for these algorithms, modern methods prefer reinforcement Learning framework as the learning part of the proposed methods. In [1], Reinforcement Learning (RL) assisted Evolutionary Algorithms (EA) have reviewed as a survey. In the paper, the usage of RL with the evolutionary algorithm classified with the usage of RL. The classification is established as integration of RL, its strategy, attribute setting and application domains. The classification has made with the direct or indirect integration, continuous or binary application, setting the objective function, reward function or action/policy setting, multi operator selection, multi population selection parameter adaptation are the main themes of these categories. However, all these categories are related to the hybrid usage of the RL or EA with each other in or instead of an operator in RL or EA.

RL and nature-inspired swarm algorithm Artificial Bee Colony (ABC) has proposed and applied to Job-shop scheduling problem [2]. In the paper the aim is to split jobs to multiple subplots so transported separately. The problem has divided into two stages and the second state RL is applied

to best scheme for the subplot, directly. In [3], hybrid flow-shop scheduling problem (DHHFSP) has utilized with Q-learning and Particle Swarm decomposition where the proposed multiobjective framework combines multi directional particle swarm global search with reinforcement learning-based local search.

In [4], tunnel deformation for urban underground safety has been investigated as engineering problem by solving the problem from finite-difference analyses using a Plastic Hardening (PH) soil model. In the paper another hybrid methodology/framework has proposed by the authors so that a hybrid machine-learning framework integrated with particle swarm optimization (PSO), convolutional neural networks (CNN), and extreme gradient boosting (XGBoost). Another application for the RL improved methods with Tabu search has presented for coloring problem in [5] where it was tested on 60 benchmark instances. In [6], an electrical problem for the shipboard has proposed and modeled this problem as shipboard multi-angle photovoltaic system with the synergistic Hybrid Evolutionary Algorithm.

A hybrid framework that integrates Quantum-Inspired Particle Swarm Optimization (QIPSO) with Deep Q-Network (DQN) reinforcement learning has been proposed and its performance has evaluated on five real-world scientific workflows, Montage, SIPHT, CyberShake, Epigenomics, and Inspiral across heterogeneous edge nodes [7]. In [8], time series prediction for six have solved with a novel proposed algorithm named as Hybrid Optimization Expert System that integrates multiple evolutionary algorithms (Arctic Puffin Optimization (APO) algorithm, Bloodsucking leech optimization algorithm (BSLO), Enterprise Development Optimization Algorithm (EDO), The Flying Lizard Optimization (FLO)) and employs BiLSTM (SJ-LSTM)a transmission mechanism, memory system, and punishment system to achieve collaborative optimization. In [9], like other studies proposes a novel multi-exemplar driven teaching learning based optimization has proposed and applied to benchmark problems CEC2014, CEC2019, and real word application benchmark problem set. A reinforcement learning-based a risk aware portfolio optimization method named Risk aware Trader has proposed as a crucial decision-making component of intelligent financial system [10]. The performance of the proposed algorithm demonstrated on numerical experiments with some public datasets, namely the Dow Jones Industrial Average and the China Securities Index.

The Q-learning method with Reinforcement learning contains potentials to be used with PSO algorithm. Unlike the previous studies it is possible to remodeled the PSO working with Reinforcement learning to select the best possible parameter sets among the action set based on diversity of the position of the population on the decision space and the improvement of the objective function on the objective space. For this reason in this research Reinforcement Learning-based Particle Swarm Optimization is proposed for single objective problems.

This paper is organized as five sections beginning with the introduction and conclusion as the final section. Section two is given for Problem statement that Reinforcement Learning, and Particle Swarm Optimization algorithms are briefly explained. In Section 3 PSO is redesigned to be used with Reinforcement Learning and in Section 4 the implementation results for this research is presented.

## 2 Problem Statement

RL, one of the machine learning methods, has been shown to be effectively used in conjunction with optimization algorithms in light of studies in the literature. These uses ensure that optimization problems are solved more efficiently and/or more accurately. However, in the literature, RL optimization algorithms are used in conjunction with other methods. RL is either used in place of one or more operators or to reduce the performance and computational cost of the optimization

algorithm. In this study, the PSO algorithm, one of the most well-known algorithms among swarm intelligence methods, will be converted into an RL framework. The algorithm will be transformed into an RL-like form. Subsequently, an assessment will be made of which section of the literature corresponds to this framework. In this respect, part of the research presented in this study offers a different perspective on the literature review.

### 2.1 Reinforcement Learning

Reinforcement Learning (RL) is a learning methodology (other machine learning methods are unsupervised learning and supervised learning) that maps the states from the environment to the actions transfer to the environment by using an assessment function called the reward. In RL the method called agent has two common input and one output. The inputs are the observations from the environment and the rewards, the output is the action. The agent contains two main operators called the policy and RL algorithm. The function takes the observations from the environment and decide an action, that function called policy. The RL algorithm is a methodology to update the policy based on the observation generated action and the reward function. Fig. 1 presents the graphical representation of the RL.

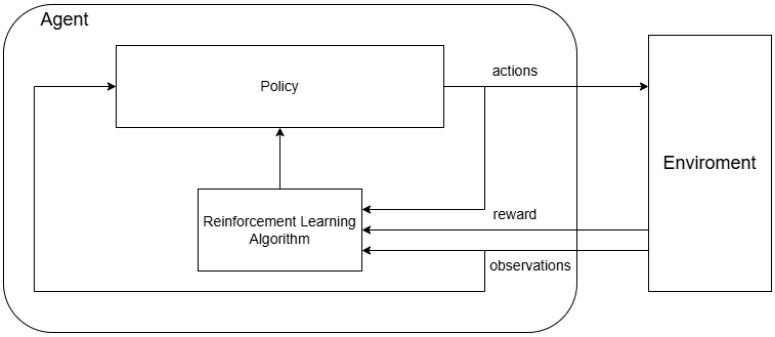



Fig. 1: General Graphical Description for Reinforcement Learning Framework.

As given in Fig. 1, the aim of the policy is to generate a n action for the desired operator by using the given observation and updates from the RL algorithm. The RL algorithm gets the action generated from policy, the reward function and the observations to update the policy to get the desired operator.

The RL methodology contains four works. These are environments, reward, policy, and training. The environment is the remaining part of the overall system except RL. This is the part that RL submits the action and gets the observations and reward. However, based on the evaluation of the environment the RL gets model-free RL and Model-based RL methods. In model free RL, the agent does not know the environment (for example dynamics of the environment) only gets the observations and policy. That means the agent do not know prior information related to the environment, therefore the agent must search the environment deeply to get a better reward. In model-based RL the agent knows (maybe partly) some parts or everything related to the environment. For example, a total map for an apart of the map for the path planning problem. Also, the environment is considered in a simulation environment of the real-life environment. The real environments present the

accurate performance form the RL however the damage to the environment is inevitable in meany engineering applications. In simulation environment is a low-cost RL application however it is not every time possible to get an accurate solution when transferred to the real word.

The reward is the numeric version of the evaluation of the action generated by the policy. It is the degree of the evaluations of how good the performance of the environment for a given action. The policy could be sparse, discrete, or continuous functions. An important part related to the selection of the reward function is to get a tradeoff between exploration and exploitation behavior.

The policy is a function that maps the observations to the actions. The policy could be direct or indirect that in direct manner the observations directly evaluated inside the policy and the indirect manner the observations are applied to the metrics and metric values are inserted to the policy. If the actions are discrete or represented by discrete states it is possible o define a look-up table. This look-up table is called Q-table that maps states/observations to actions. Instead of continuous function and Q-table, neural networks can be used as policy that is the most preferred approach in AI applications.

The policy structured into three different approaches. These are policy function based, value function based and actor-critic approaches. The policy function-based learning is the entire learning method for a single structure is the entire policy like neural network. For example, the output of the neural network are the discrete events on a system. Similarly, policy gradient methods that gradient of the rewards and/or the policy uses to decide the direction of the better reward value. Value function-based policy is a system that not generate action solely. Instead, it generated some value and based on that value the action is decided. for example, the Q-table and the bellman equation to update the entities of the Q-table can be given.

Actor-critic policy and learning is a continuous network so that the actor generates the action and the critic predicts the reward value. In general, the neural network is the actor because it directly generated the actions. As critic neural network also utilize to predict the reward value. Similarly, both networks trained with the observations.

### 2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) was initially proposed by Kenndy and Eberhart in 1995 [11]. The main motivation of the algorithm is to integrate the nature inspired swarm intelligence to the search algorithm that formed an optimization algorithm. The algorithm is depended on the movement of the particles on the objective space based on the movements of the members in an animal flock or swarm. The change of positions of the members in a flock becomes solution candidates on the population-based PSO algorithm. Since the PSO is relatively simple to implement and has relatively low number of parameters that needed to adjusted to get a desired performance; are two reasons to get attention from the engineering problems as a stochastic optimization methodology. The PSO algorithm contains position and velocity update rules as given in Eq. 1.

$$\begin{aligned} v(k+1) &= wv(k) + c_1 r_1 \left(pb - x(k)\right) + c_2 r_2 \left(gb - x(k)\right) \\ x(k+1) &= x(k) + v(k+1) \end{aligned} \tag{1}$$

where $v$ is the velocity, $x$ is the position, $k$ is the iteration index, $pb$ is the personal best from the beginning of the iteration to the $k$th iteration, $w$, $c_1$ and $c_2$ are the control parameters with two random number generators $r_1$ and $r_2$ within $[0, 1]$.

The performance of the PSO algorithm have been improving through more than thirty years after the proposing the PSO in terms of exploration and exploitation. Algorithm 1 gives the pseudo-code for the PSO algorithm. Since PSO is a population-based algorithm, it begins with the initial definition of the parameters and position. Next the algorithm begins with global and personal best member detection and velocity update and position update rules. This procedure continues until the maximum number of iterations is reached. This algorithm suffers from the local search with local optimum, early convergence and lower accuracy. Therefore, many variants have proposed to improve the performance. The impact of some of the main improvements will be discussed after the redesigning of PSO in the next sub-section.

**Algorithm 1** Particle Swarm Optimization (PSO)

Select random number generator, number of members in population $N$, maximum number of iterations $M$ and parameters $w$, $c_1$ and $c_2$
**for** particle $i$ **do**
  Define initial position for each particle $x(0)$
  Calculate Objective function $f(x(0))$
**end for**
**for** $k = 1$ increase k $k = k + 1$ **do**
  Update Personal best $pb$
  Update Global best $gb$
  Update Velocity

$$v(k+1) = wv(k) + c_1 r_1 \left(pb - x(k)\right) + c_2 r_2 \left(gb - x(k)\right) \tag{2}$$

  Update Position

$$x(k+1) = x(k) + v(k+1) \tag{3}$$

  Calculate Objective function $f(x(k))$
**end for**

## 3 Redesigning Particle Swarm Optimization Algorithm

The Algorithm 1 given for the PSO presents the flow of the algorithm in a sequential manner. The PSO algorithm mainly contains the update rule with the parameters the objective function. The nature of this algorithm suitable to map the sequential algorithm to the RL framework. Fig. 2 graphically demonstrate the PSO algorithm in the form of RL framework.

As given in the figure, RL framework has two main blocks that are policy and the learning algorithm with reward function. It is clear to state that the reward function easily maps to the reward of the optimization problem. The action is defined as the velocity and the velocity update rule is the policy of this structure. Therefore, the position is the observation from the environment. The position also the solution candidate that is evaluated from the objective function also called the reward function.

The learning algorithm has four parameters that have to be set. Luckily, conventional PSO has constant values for $c_1$, and $c_2$. The personal best value and the global best value calculated from the population by using the min function by using the reward function. The other parameter $w$ is the weighting multiplier set to be decreasing with the iteration.

In the following two sub-sections the proposed improvements for the PSO are evaluated with respect to the policy and learning algorithms. Therefore, in this research the improvements on the

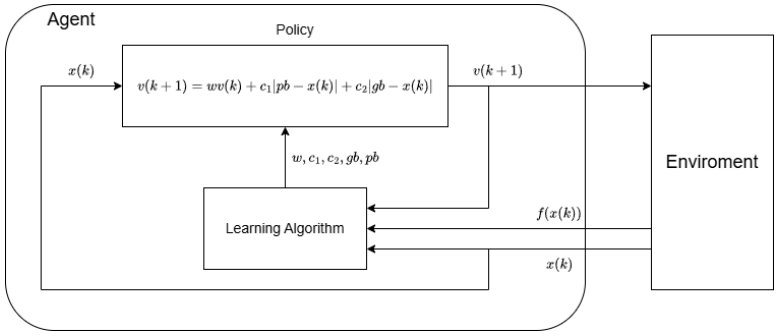



Fig. 2: General Graphical Description for the Particle Swarm Optimization Algorithm formed as Reinforcement Learning Framework.

algorithm categorized into two sets and it is possible to categorized other variants of the PSO. However, to maintain the content of the research only the well-known methods are selected and discussed.

### 3.1 Learning Algorithms

The learning algorithms are sub-routines that adjust the parameters of the PSO algorithm. Therefore, the learning is to set these values based on the observation actions and reward function. As conventional way, the learning is the decision method for the parameters related to the policy of the PSO variant. In general, the intelligence methods to decide these parameters corresponding to the learning algorithm. In general, three main groups are well known which are parameter update, inertia weight and constriction factor. There are other methods based on policy of the algorithm.

Parameter Update: Clerc's study [12] has the possible selection of the parameters $c_1$ and $c_2$ so that it is possible to considered the corresponding values of these two parameters gives the decision between cognitive-dominant ($c_1 > c_2$); social-dominant ($c_1 < c_2$) and balanced behavior ($c_1 = c_2$) (in [12] it was stated that $c_1 = c_2 = 2.05$) between personal best and global best tendency between particles [13].There are many possible selection for these values like in [14] $c_1 = c_2 = 1.49$ parameter set was selected.

Inertia Weight: In [15], a weighting parameter is defined for the PSO so that as given in Eq.1 the weight multiplied with the previous velocity. This parameter uses to balance exploration and exploitation. The study of [16] the suggestion for the weight is in the range between 0.5 and 0.8. The adaptive exponential linear and nonlinear variations of the weight based on iteration or rank can be selected as a weighting variant. In [13] these methods have given as a summarized table, and also given in [17].

Constriction Factor: The parameter is given to multiply the overall velocity update rule (without the weight) to increase convergence and stability of the PSO algorithm suggested in [12]. The common value for this parameter is 0.729 [12].

### 3.2 Policy

After redesigning the PSO as discussed in the previous section, the PSO algorithm have divided into two parts, the policy and the learning. In this subsection the policy will be discussed. The policy

is the function that generates the action therefore for the PSO as explained it is corresponding to the velocity update formulations for each variant.

Velocity Clamp: The first and easy part is the restrict the obtained velocity. This action corresponds to the actor-critic manner so that the critic related to the generated action is made by comparing the maximum and minimum velocity values. The restriction to the velocity is a common methodology to restrain the position of the population. The maximum and minimum velocity helps the algorithm to remains in a given position range [18]. It is possible to change this maximum and minimum values at each iteration.

Velocity Update Rules: The velocity update rules changed, improved, merged with other operator to get more accurate solution with a smaller number of iterations. As an example in [19] two mutation operators merged with PSO to update velocity. Similarly in [20], differential mutation operator with social learning mechanism utilized with PSO. In [21], the velocity update rule is changed by introducing some novel parameters like in [22] where a third summation is added to the velocity update rule with a new parameter $c_3$ and a direction value.

### 3.3 Proposed RL-PSO Algorithm

The proposed RL-PSO algorithm begins with the random initialization of the positions and velocity. Then the observations are determined. Observation space is selected as six states. The state is determined from the global best position $s(t) = f(g^{best}(t) - g^{best}(t-1))$. Next the action space is decided. The action space contains three parameters $w$, $c_1$ and $c_2$. These parameters took the values inside $[0.1, 2]$ with 10 discrete and equally divided values. Next based on the selected parameter the position and the velocity of the members are updated.

$$v_{i,d}(t+1) = w \cdot v_{i,d}(t) + c_1 r_1 \big(p_{i,d}^{best} - x_{i,d}(t)\big) + c_2 r_2 \big(g_d^{best} - x_{i,d}(t)\big), \tag{4}$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1), \tag{5}$$

Unlike previous discussion, in this RL-PSO the reward function is defined as $r(t) = g^{best}(t-1) - g^{best}(t)$ which means the reward is the difference between global best objective function values at the current and the previous iteration. Next, the Q-table is updated and new state is determined. In RL-PSO algorithm the action parameters are determined from Q-learning agent.

The update of the Q-table is made by Bellman function like the conventions Q-function training methodology given as

$$Q(s(t), a(t)) \leftarrow Q(s(t), a(t)) + \alpha \Big[ r(t) + \gamma \max_{a' \in A} Q(s(t+1), a') - Q(s(t), a(t)) \Big], \tag{6}$$

where $\alpha \in (0, 1)$ is the learning rate, and $\gamma \in (0, 1)$ is the discount factor. The policy selection is made from $\varepsilon$-greedy policy.

Unlike conventional way that the velocity update rule is the policy; in RL method the policy is the $\varepsilon$-greedy and the training rule is given in Eq. 6. The next section shows the implementation of RL-PSO algorithm and comparison between PSO algorithm is proposed.

In this framework, the difference between global best objective value of the current iteration and the previous iteration is selected as the improvement. In addition the average of the standard deviation of the members in the population on the decision space is the diversity of the population. Based on improvement and diversity of the solutions the states are selected. The Q-table contains

6 states and 8000 actions. the actions are formed from three parameters $w$, $c_1$ and $c_2$ and their 20 different values. The Q-table values are updated with Bellman functions and based on the current state and the Q values the action is selected among the 8000 actions.

Table 1: State Transition Logic Based on Improvement and Diversity

| **State** | **Improvement** | **Diversity** | **Interpretation** |
|---|---|---|---|
| 1 | $> 10^{-2}$ | $> 1$ | Fast Growth |
| 2 | $> 10^{-4}$ | $> 1$ | Approximate Steady State |
| 3 | $> 0$ | $\leq 1$ | Steady state |
| 4 | $\leq 0$ | $> 1$ | Wide Stagnation |
| 5 | $\leq 0$ | $> 0.1$ | Narrow Stagnation |
| 6 | Else | $\leq 0.1$ | End |

Table 1 gives the state ids and their approximate interpretations with respect to the ranges of the improvement and diversity. As the improvements are increased with the diversity of the populations that means searching a relatively large area of the objective space (exploration) and as the improvement reduces, became smaller with a more crowded populations means the solution becomes converging therefore the algorithm is approximately in the steady state.

## 4 Implementation

In this section, the brief information related to the benchmark problems and their properties will be explained. Next, the implementation results and their convergence will be reported as the results and discussions.

### 4.1 Benchmark Problems

In this research to demonstrate the performance of the proposed RL-PSO algorithm explained in the previous section is applied to the CEC 2017 benchmark problem set defined in [23]. The benchmark set contains 30 problems however problem 2 is deleted after the definition of the problems (in total there are 29 problems) where the codes for the Matlab file can be found in [24]. The search space for all these benchmark problems are in $[-100, 100]^D$ where $D$ is the dimension of the search space which is equal to 30 for all problems. The problems are categorized into four groups, problem 1 and 2 are unimodal Functions, Problem 3-9 are simple multimodal functions, problems 10-19 are hybrid functions and problems 20 - 29 are composition functions. First 9 problems are actually the shifted and rotated of the conventional benchmark problems like Rosenbrock and Rastrigin's Functions. Problem 10-19 are hybrid problems so that the linear combination of the different basic functions is weighted and summed as sub components of these benchmark problems. Finally, there are composition functions where these functions are like hybrid functions with weights, biases and shifted functions. Unlike hybrid problems, in composition problems all basic functions have been used as shifted and rotated. In [23], the mathematical expressions and their search spaces are graphically demonstrated. The optimum values for each benchmark problems are equal to problem ID multiply by 100. For example, the optimum solution for the final problem is 2900.

### 4.2 Results and Discussions

In this research Reinforcement Learning is used to determine the best $w$, $c_1$ and $c_2$ parameters. The obtained algorithm named as RL-PSO algorithm. The performance of this algorithm will be compared with PSO meta-algorithm under same conditions. All algorithms have 100 members in their population 30 and iteration is 200. The implementations are repeated 30 times and the statistics of these results are reported numerically. In addition, Wilcox statistical test is provided to show the significant of the results. for the PSO algorithm the parameters are selected as $w = 0.7$, $c_1 = 2$, and $c_2 = 2$. The implementations reevaluated and Table 2 gives the results numerically and Fig 4 shows the average convergence plots for all of the benchmark problems.

Table 2 gives the statistical properties of the 30 independent run with mean and standard deviation. In addition, Wilcoxon significant test results are added and it is numerically showed that the proposed algorithm is better than PSO for 12 problems and almost same for 11 other problems. Also, for the rest of the problems PSO gives better objective values, performance. For unimodal and simple multimodal problems, the proposed method presents better results however for composition and hybrid problems the performance of PSO and RL-PSO gives almost same performance.

Figure 4 gives the selected actions for each problems. Each plot in Figure 4 gives for each benchmark problems. The plots are given for each iteration that means for each iteration a different action is selected based on the Q-value. These plots are among the 30 independent values the most frequently selected action for a given iteration. As a general perspective, it can be observed from the figure that as the iteration increased in number the actions becomes more converged value for a better performance cases.

## 5 Conclusion

In this research the PSO algorithm parameters $w$ $c_1$ and $c_2$ determined with the aid from reinforcement learning so that the Q-table is generated. From the table the best parameters are obtained for PSO algorithm under CEC 2017 test problems. The results are numerically and graphically compared with each other. The results showed that RL-PSO presents mostly better results. However, since PSO is relatively simple nature and easy to implement and has lower computational necessities, it is possible to comment that even RL-PSO gives better results the desired computational resources is relatively highly for this performance. As future study, learning-based algorithms are more deeply discussed to get a better single and multi-objective problems.

## References

1. Yanjie Song, Yutong Wu, Yangyang Guo, Ran Yan, Ponnuthurai Nagaratnam Suganthan, Yue Zhang, Witold Pedrycz, Swagatam Das, Rammohan Mallipeddi, Oladayo Solomon Ajani, Qiang Feng: Reinforcement learning-assisted evolutionary algorithm: A survey and research opportunities, Swarm and Evolutionary Computation, **86**(2024), 101517, (2024), https://doi.org/10.1016/j.swevo.2024.101517.
2. Yibing Li, Cheng Liao, Lei Wang, Yu Xiao, Yan Cao, Shunsheng Guo: A reinforcement learning-artificial bee colony algorithm for flexible job-shop scheduling problem with lot streaming, Applied Soft Computing, **146**(2023), 110658, (2023) https://doi.org/10.1016/j.asoc.2023.110658.
3. Wenqiang Zhang, Sen Yang, Mingzhe Li, Yashuang Mu, Guohui Zhang, Mitsuo Gen, Peng Li: A hybrid multiobjective particle swarm optimization with Q-learning-driven local search for distributed heterogeneous hybrid flow-shop scheduling problem with worker–machine–environment collaboration. Journal of Manufacturing Systems,**85**(2026), 1–21, (2026) https://doi.org/10.1016/j.jmsy.2025.12.024.

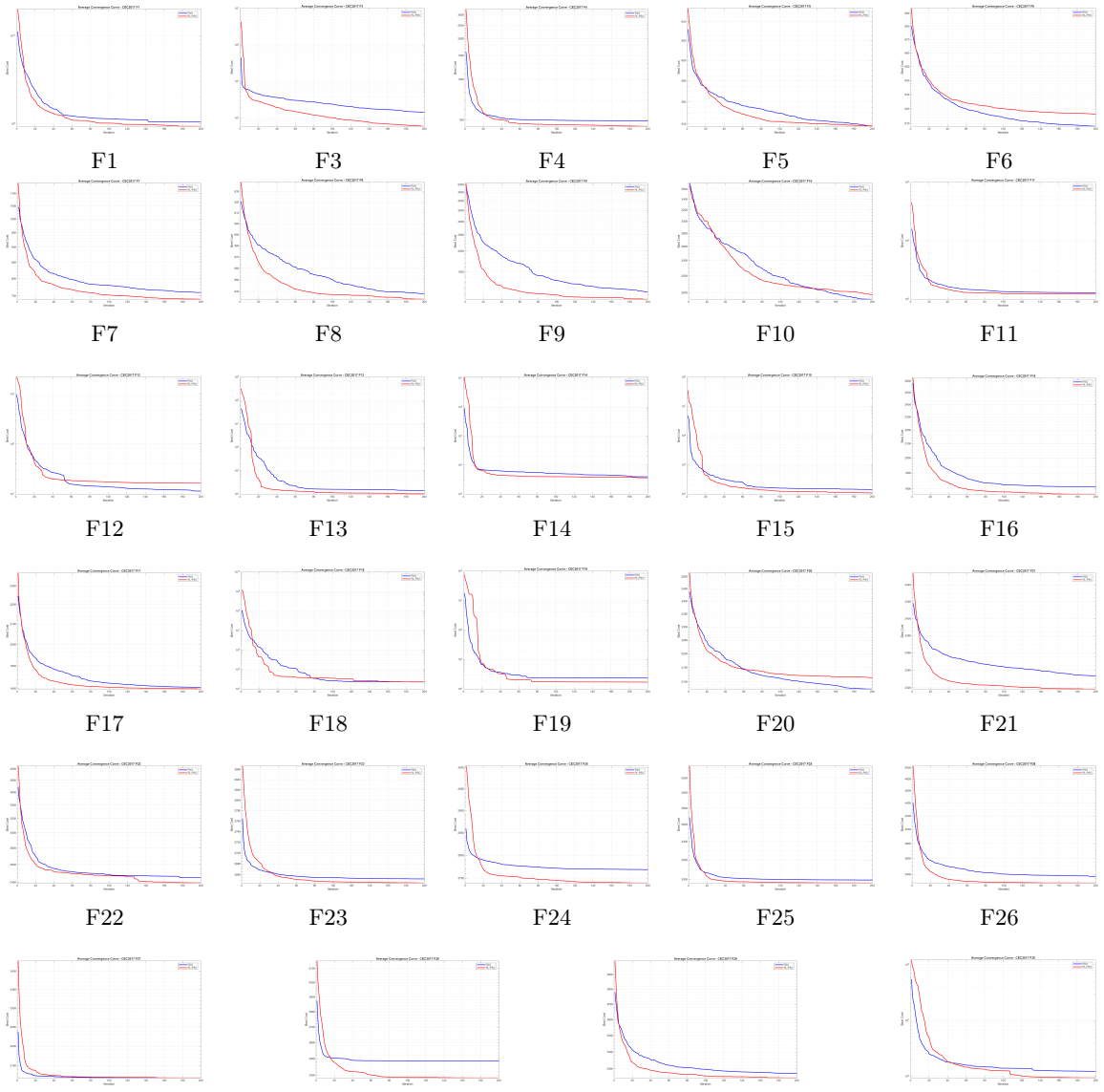



Fig. 3: Average convergence curves for 29 CEC 2017 functions

Table 2: Statistical properties of the optimization algorithms based on their mean and standard deviation of the objective values.

| Prob. ID | PSO | RL-PSO |
|---|---|---|
| 1 | 01.37e+09 (01.21e+09) | 03.55e+08 (06.63e+08) $-$ |
| 3 | 13.78e+02 (95.31e+02) | 50.08e+02 (70.83e+02) $-$ |
| 4 | 49.05e+01 (11.89e+01) | 44.03e+01 (54.74e+00) $-$ |
| 5 | 53.83e+01 (10.594e+00) | 54.20e+01 (18.81e+00) $+$ |
| 6 | 60.88e+01 (06.86e+00) | 61.25e+01 (10.34e+00) $+$ |
| 7 | 75.89e+01 (14.51e+00) | 75.64e+01 (42.93e+00) $-$ |
| 8 | 83.75e+01 (10.67e+00) | 83.86e+01 (18.83e+00) $\approx$ |
| 9 | 10.45e+02 (23.14e+01) | 12.12e+02 (45.81e+01) $+$ |
| 10 | 19.32e+02 (30.05e+01) | 19.45e+02 (39.62e+01) $\approx$ |
| 11 | 13.29e+02 (27.27e+01) | 13.06e+02 (55.25e+01) $\approx$ |
| 12 | 01.77e+07 (05.35e+07) | 01.35e+06 (02.69e+06) $-$ |
| 13 | 14.56e+03 (14.22e+03) | 93.24e+02 (10.09e+03) $-$ |
| 14 | 38.10e+02 (50.81e+02) | 55.97e+02 (71.67e+02) $+$ |
| 15 | 13.00e+03 (19.94e+03) | 10.67e+03 (11.97e+03) $-$ |
| 16 | 17.82e+02 (12.55e+01) | 18.16e+02 (16.39e+01) $\approx$ |
| 17 | 18.10e+02 (61.01e+00) | 17.95e+02 (62.33e+00) $\approx$ |
| 18 | 30.14e+03 (13.89e+03) | 42.39e+02 (74.82e+03) $+$ |
| 19 | 33.74e+03 (56.59e+03) | 15.10e+02 (21.76e+03) $-$ |
| 20 | 20.84e+02 (54.72e+00) | 21.37e+02 (67.32e+00) $+$ |
| 21 | 23.39e+02 (09.55e+00) | 23.29e+02 (37.83e+00) $\approx$ |
| 22 | 23.75e+02 (19.73e+01) | 23.58e+02 (14.98e+01) $-$ |
| 23 | 26.50e+02 (23.15e+00) | 26.51e+02 (27.38e+00) $\approx$ |
| 24 | 27.72e+02 (42.35e+00) | 27.73e+02 (61.54e+00) $\approx$ |
| 25 | 29.84e+02 (60.02e+00) | 29.81e+02 (97.89e+00) $\approx$ |
| 26 | 33.22e+02 (44.45e+01) | 31.45e+02 (34.58e+01) $-$ |
| 27 | 31.07e+02 (11.31e+00) | 31.25e+02 (34.83e+00) $\approx$ |
| 28 | 34.44e+02 (14.86e+01) | 33.54e+02 (12.61e+01) $-$ |
| 29 | 32.87e+02 (73.23e+00) | 32.74e+02 (81.30e+00) $\approx$ |
| 30 | 09.34e+05 (06.78e+05) | 05.08e+05 (05.72e+05) $-$ |
| $+/-/\approx$ | | 6/12/11 |

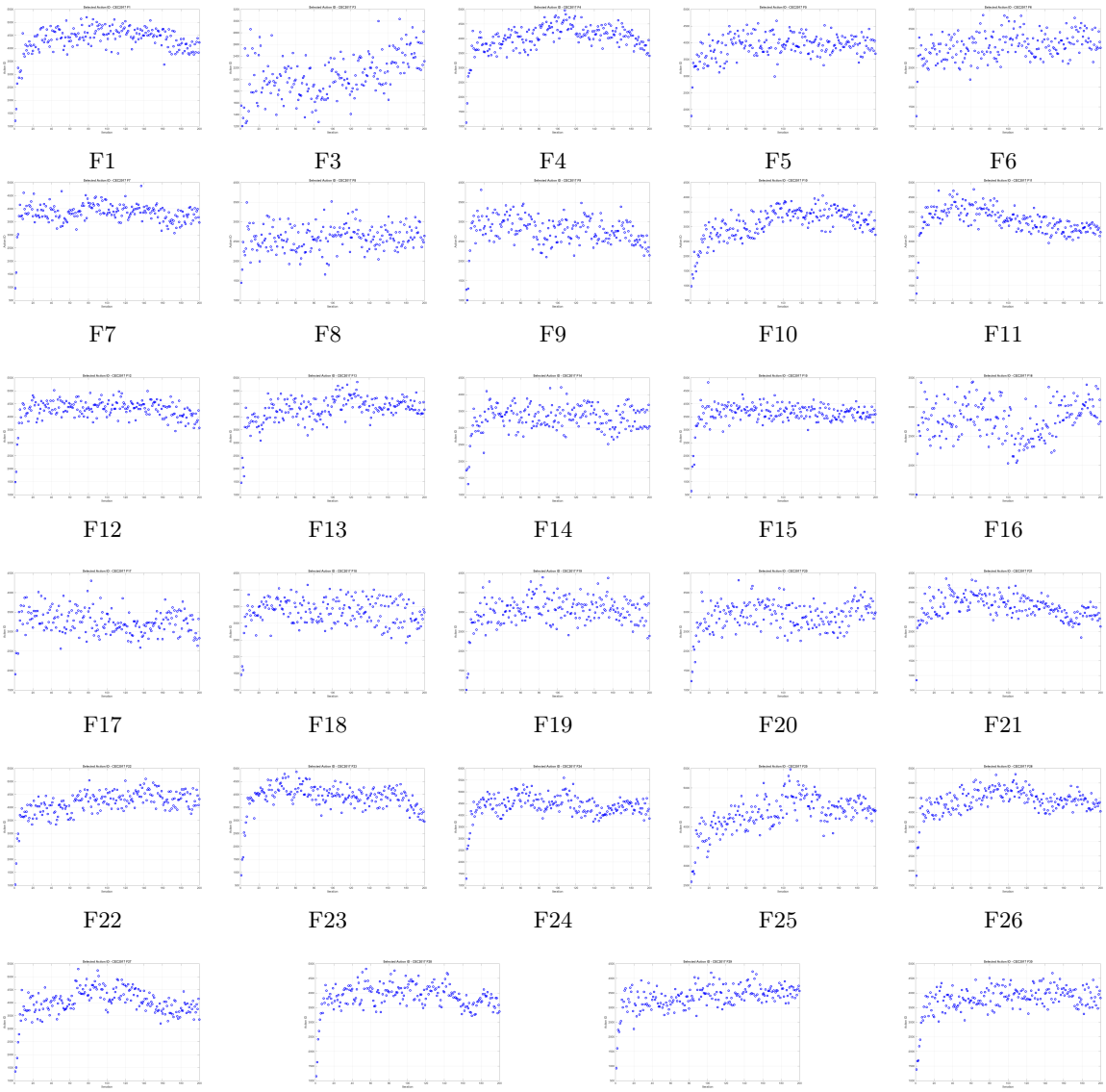



Fig. 4: Average selected Action ID for for 29 CEC 2017 functions

4. Siyu Yin, Zheng Yang, Kunpeng Cao, Ben Wu, Siau Chen Chian: Plastic-hardening constitutive model-based hybrid machine learning framework for three-dimensional tunnel deformation prediction, Tunnelling and Underground Space Technology. **171**(2026), 107415, (2026) https://doi.org/10.1016/j.tust.2025.107415.
5. Zhe Sun, Una Benlic, Mingjie Li, Qinghua Wu: Reinforcement learning based tabu search for the minimum load coloring problem, Computers and Operations Research, **143**(2022), 105745, (2022) https://doi.org/10.1016/j.cor.2022.105745.
6. Qilin Xiang, Lijie Xu, Chengqing Yuan,: A reinforcement learning-based Synergistic Hybrid Evolutionary Algorithm for multi-angle shipboard photovoltaic system MPPT under dynamic navigational shading. Renewable Energy,**260**(2026), 125193, https://doi.org/10.1016/j.renene.2026.125193.
7. Megha Sharma, Abhinav Tomar: QERLO An intelligent Quantum-Enhanced Reinforcement Learning framework for task Offloading in IoT networks. Advanced Engineering Informatics, **70**(2026), 104096, https://doi.org/10.1016/j.aei.2025.104096.
8. Wei, P., Fan, C., Yang, X. et al. HOES: an efficient multi-evolutionary expert system for deep learning model optimization in time series prediction. Sci Rep 16, 527 (2026). https://doi.org/10.1038/s41598-025-30014-4
9. Sharma, A., Lim, W.H., Berghout, T. et al. Process Innovation via Adaptive Multi-exemplar Driven Optimization: Enhancing Engineering Design and Global Optimization Performance. Int J Comput Intell Syst 19, 22 (2026). https://doi.org/10.1007/s44196-025-01086-x
10. Yang, M., Wang, J. and Hu, Y. RiskawareTrader A Reinforcement Learning based Portfolio Optimization for Risk Averter. Int J Comput Intell Syst 19, 25 (2026). https://doi.org/10.1007/s44196-025-01094-x
11. J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, 1995, pp. 1942-1948 vol.4, doi: 10.1109/ICNN.1995.488968.
12. M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," IEEE Transactions on Evolutionary Computation, vol. 6, no. 1, pp. 58-73, Feb. 2002, doi: 10.1109/4235.985692.
13. Rashmi Sharma, Jagjit Singh Matharu, Kulwinder Singh Parmar: A survey on Particle Swarm Optimization: Evolution, adaptations and practical implementations, Applied Soft Computing, Volume 186, Part A, 2026, 114016, https://doi.org/10.1016/j.asoc.2025.114016.
14. A. Ratnaweera, S. K. Halgamuge and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," in IEEE Transactions on Evolutionary Computation, vol. 8, no. 3, pp. 240-255, June 2004, doi: 10.1109/TEVC.2004.826071.
15. Y. Shi and R. Eberhart, "A modified particle swarm optimizer," 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), Anchorage, AK, USA, 1998, pp. 69-73, doi: 10.1109/ICEC.1998.699146.
16. Lin Xueyan and Xu Zheng, "Swarm size and inertia weight selection of Particle Swarm Optimizer in system identification," 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), Harbin, 2015, pp. 1554-1556, doi: 10.1109/ICCSNT.2015.7491026.
17. T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh and S. Mirjalili, "Particle Swarm Optimization: A Comprehensive Survey," in IEEE Access, vol. 10, pp. 10031-10061, 2022, doi: 10.1109/ACCESS.2022.3142859.
18. Russ Eberhart, Pat Simpson, and Roy Dobbins. 1996. Computational intelligence PC tools. Academic Press Professional, Inc., USA.
19. Yonggang Chen, Lixiang Li, Haipeng Peng, Jinghua Xiao, Yixian Yang, Yuhui Shi: Particle swarm optimizer with two differential mutation, Applied Soft Computing, Volume 61, 2017, 314-330, https://doi.org/10.1016/j.asoc.2017.07.020.
20. Xinming Zhang, Xia Wang, Qiang Kang, Jinfeng Cheng: Differential mutation and novel social learning particle swarm optimization algorithm, Information Sciences, Volume 480, 2019, 109-129, https://doi.org/10.1016/j.ins.2018.12.030.

21. Davoud Sedighizadeh, Ellips Masehian, Mostafa Sedighizadeh, Hossein Akbaripour: GEPSO: A new generalized particle swarm optimization algorithm, Mathematics and Computers in Simulation, Volume 179, 2021, 194-212, https://doi.org/10.1016/j.matcom.2020.08.013.
22. Abdullah, M.N. and Bakar, Ab Halim Abu and Rahim, N.A. and Mokhlis, Hazlie and Illias, Hazlee Azil and Jamian, J.J. (2014) Modified Particle Swarm Optimization with Time Varying Acceleration Coefficients for Economic Load Dispatch with Generator Constraints. Journal of Electrical Engineering Technology, 9 (1). pp. 15-26. ISSN 1975-010
23. N. H. Awad, M. Z. Ali, P. N. Suganthan, J. J. Liang and B. Y. Qu, "Problem Definitions and Evaluation Criteria for the CEC 2017 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization" Technical Report, Nanyang Technological University, Singapore And Jordan University of Science and Technology, Jordan And Zhengzhou University, Zhengzhou China, 2016.
24. P-N-Suganthan, CEC2017-BounContrained, Available in https://github.com/P-N-Suganthan/CEC2017-BoundContrained

# Application of the RRT Algorithm to Solve Control Theory Problems

Kira Dmitrieva[1] and Majid Abbasov[1][0000-0003-1484-4733]

[1] Universitetskaya Emb., 13B, St Petersburg 199034, Russia

**Abstract.** This paper presents a modified Rapidly-exploring Random Tree (RRT) algorithm for solving optimal control problems in nonlinear dynamical systems. The key innovation lies in integration of system dynamics direct into the tree expansion process. Unlike the classical RRT, which samples points in the configuration or state space, the proposed approach samples controls from a mixed continuous-discrete probability distribution. Each sampled control is applied to the nearest node in the tree, and the resulting child state is computed by numerically integrating the system's ordinary differential equations over a fixed time step. To explicitly track and minimize the objective functional, the state vector is augmented with the accumulated cost, forming an extended state space. This framework transforms the RRT from a path-planning tool into a stochastic optimizer capable of approaching a control trajectory that satisfies the necessary conditions for optimality while respecting state and control constraints. The method is validated through numerical simulation, demonstrating its effectiveness in generating feasible and near-optimal control policies for complex dynamical systems.

**Keywords:** optimal control, graph tree, fast growing graph trees, control theory.

## 1 Introduction

The Rapidly-exploring Random Tree (RRT) algorithm was introduced by LaValle (1998) as a randomized data structure designed for path planning problems with nonholonomic constraints [1]. Unlike grid-based methods, RRT efficiently explores high-dimensional state spaces by incrementally building a tree toward randomly sampled nodes. A major advancement came with RRT-Connect, developed by Kuffner and LaValle (1999), which grows two trees simultaneously—one from the start and one from the goal—significantly improving convergence speed for single-query problems [2]. While the original RRT guarantees probabilistic completeness, it lacks asymptotic optimality. This limitation was addressed by Karaman and Frazzoli (2011), who proposed RRT*, a variant that introduces rewiring of the tree to ensure almost-sure convergence toward the optimal path as the number of iterations increases [3]. Following the introduction of RRT*, numerous modifications have focused on accelerating convergence and improving solution quality. A key challenge in applying RRT to control problems is the incorporation of system dynamics and cost functions.

The classical RRT framework assumes geometric steering, which is insufficient for systems governed by ordinary differential equations (ODEs). LaValle and Kuffner laid the groundwork for kinodynamic planning by extending RRT to systems with differential constraints. Later it was demonstrated how RRTs can be adapted for nonlinear control and hybrid systems, replacing geometric distance metrics with cost-to-go heuristics and integrating local planners for dynamic feasibility. However, despite significant progress, RRT-based algorithms face persistent theoretical and practical challenges.

## 2 Problem statement

Consider search for optimal control problem with parameters:

$$\begin{cases} \ddot{x} + \omega^2 x = u, t \in [0; T] \\ x(0) = x_0 \\ x(T) = x_* \\ x(t) \in X, t \in [0; T] \\ u(t) \in U, t \in [0; T] \end{cases}$$

$x(t) \in \mathbb{R}^n$- state vector
$u(t) \in \mathbb{R}^m$- control vector
$X \subseteq \mathbb{R}^n$- state space
$U = [u_{min}; u_{max}]$- vector of permissible controls
$\omega$ – constant parameter, angular frequency
$T > 0$ – control horizon

The equation can be rewritten as a system: $\begin{cases} \dot{x_1} = x_2 \\ x_2 = -\omega^2 x_1 + u \end{cases}$

The task is to transfer the system from point $x_0$ to point $x_*$ in time $T > 0$ minimizing the functional $J$:

$$J = \frac{1}{2}\int_0^T u^2(t)dt \quad (1)$$

Consider the system $z = \begin{pmatrix} x \\ s \end{pmatrix} : \dot{s}(t) = \frac{u^2}{2}$

Then the system can be rewritten as:

$$\dot{z}(t) = \begin{pmatrix} x_2 \\ -\omega^2 x_1 + u \\ \frac{u^2}{2} \end{pmatrix} \quad (2)$$

That is, the problem reduced it to the finding a trajectory in an extended space: $\forall x \in S: S = \{z = \begin{pmatrix} x^* \\ s \end{pmatrix} : s \in \mathbb{R}\}$ with minimizing s(T). Thus, the problem is reduced to finding a trajectory in an extended state space minimizing the terminal cost. The RRT algorithm is based on probabilistic exploration of space and requires stochastic control expansion. A random sample of controls from probability distributions is used to construct a tree of reachable states:

$$U\sim \begin{cases} DUnif(\{x \in \mathbb{Z}: x \in [u_{min}; u_{max}]\}), with\ p\ probability \\ Unif([u_{min}; u_{max}]), with\ 1-p\ probability \end{cases} \quad (3)$$

Stochastic control selection guarantees that the tree will reach the target state with probability 1 as the number of iterations approaches infinity. Random sampling allows the algorithm to avoid local traps in the control space. For complex control systems with hidden spatial geometry, stochastic exploration is often more efficient than deterministic grid methods. With a sufficient number of iterations, the resulting control asymptotically approaches a control satisfying the necessary optimality conditions.

Such hybrid distribution is a heuristic. We assume that discrete part gives “jumps” to control space, giving tree more options and directions to grow, avoiding local “traps”. Continuous part helps with smoothing the found trajectories, locally improving found solutions and making them more precise. Parameter p controls the balance between exploration and refinement. The larger p is, the more continuous samples there are, the better the local improvement, and the worse the global coverage. The smaller p is, the stronger the discrete phase, the faster the coverage of rough directions, but the slower the fine tuning. A practical scheme can be described this way: start with a small pp (almost purely discrete mode) and gradually increase pp as the tree grows, in order to quickly cover the space first and then further optimise the found trajectories. This will be analogous to coarse-to-fine in optimisation tasks.

# 3 Algorithm

## 3.1 Algorithm modification description

The original RRT algorithm builds a tree in the state space by sampling random points, finding the nearest node in the tree, and extending it toward the sample via straight-line motion (or a local steering function) over a fixed distance. This process does not consider system dynamics or path costs.

Our modified code incorporates system dynamics by defining an ODE for state evolution. The state is numerically integrated over a fixed time step dt. Each state in the tree is augmented with its accumulated cost, forming an extended state vector [x, s]. This transforms the tree into a cost-aware search structure that explicitly tracks and minimizes the total cost to reach the goal. While extensions like RRT* introduce rewiring for asymptotic optimality, they typically lack this explicit integration of cost propagation within the dynamics.

New nodes are generated not by extending toward a random state, but by applying randomly sampled controls to the nearest tree node. These controls are drawn from a mixed continuous-discrete distribution. The resulting child state is computed by numerically integrating the system ODE over dt under the applied control. This approach ensures that all generated trajectories are dynamically feasible, as they respect the system dynamics exactly. The new state is then validated against state and control constraints, as well as dynamic feasibility checks. Similar to the standard RRT, the proposed method does not guarantee that the found trajectory is optimal. Therefore, claims regarding asymptotic optimality are not made.

### 3.2 Algorithm code description

The algorithm builds a tree of dynamically feasible trajectories by repeatedly selecting a random state (for nearest neighbor guidance), picking a random control from a mixed distribution, integrating the dynamics forward, and adding the resulting state to the tree if valid. The augmented state with accumulated cost enables direct comparison of trajectory quality when multiple paths reach the goal.

The algorithm takes the following parameters as input:

1. $x \in [-2; 2] \times [-2; 2]$
2. $u \in U$, where $U \sim \begin{cases} DUnif(\{x \in \mathbb{Z}: x \in [u_{min}; u_{max}]\}), with\ p\ probability \\ Unif([u_{min}; u_{max}]), with\ 1 - p\ probability \end{cases}$
3. $T = 100$c
4. $\omega = 1.0$
5. $Integration\ timestep\ dt$

## Pseudocode

**Algorithm 1** RRT for dynamic system

**class** `RRT`
  **input**:
    `dynamics`: function $f(x, u) \to \frac{dx}{dt}$
    `cost_rate`: function $c(x, u) \to$ cost
    `state_bounds`: $(x_{\min}, x_{\max})$ — state space bounds
    `control_bounds`: $(u_{\min}, u_{\max})$ — integer control space bounds
    `prob_contr`: probability of selecting a continuous control
    `goal_tol`: goal reaching tolerance
    $dt$: integration time step
    $T_{\max}$: time horizon
  **initialize**:
    `goal_tol`, $dt$, $T_{\max}$
    `n_state` ⟵ dimension of the state space
    `max_iter` ⟵ $\lfloor \frac{T_{\max}}{dt} \rfloor$
  **method** `random_control()`:
    `u_rand` ⟵ random number from $U[0, 1]$
    **if** `u_rand` $<$ `prob_contr`:
      **return** random control sampled from $U[u_{\min}, u_{\max}]$
    **else**:
      **return** random integer control sampled from discrete uniform distribution on $\{u_{\min}, \ldots, u_{\max}\}$
  **method** `integrate`$(z_{\text{near}}, u)$:
    // $z_{\text{near}} = (x_{\text{near}}, s_{\text{near}})$, $s$ — accumulated cost
    **define** ODE:

$$\begin{cases} \frac{dx}{dt} = \texttt{dynamics}(x, u) \\ \frac{ds}{dt} = \texttt{cost_rate}(x, u) \end{cases}$$

    **solve** ODE from $t = 0$ to $t = dt$ with initial conditions $z_0 = (x_{\text{near}}, s_{\text{near}})$
    **if** integration succeeds:
      **return** $z_{\text{new}} = (x_{\text{new}}, s_{\text{new}})$
    **else**:
      **return** *null*
  **method** `is_state_valid`$(x)$:
    **return** *true* **if** $x \in [x_{\min}, x_{\max}]$ **else** *false*
  **method** `goal_reached`$(x, x_{\text{goal}})$:
    **return** *true* **if** $\|x - x_{\text{goal}}\| \leqslant$ `goal_tol` **else** *false*
  **method** `plan`$(x_0, x_{\text{goal}})$:
    $z_{\text{start}}$ ⟵ $(x_0, 0)$
    **initialize** `tree`:
      `states` ⟵ $[z_{\text{start}}]$
      `controls` ⟵ $[null]$
      `parents` ⟵ $[-1]$
    `best_cost` ⟵ $+\infty$
    `best_path` ⟵ *null*
    `best_controls` ⟵ *null*
    **for** $i$ **from** 0 **to** `max_iter` $- 1$:
      $x_{\text{rand}}$ ⟵ random node in $[x_{\min}, x_{\max}]$
      `idx` ⟵ index in `tree.states` which minimizes $\|x_{\text{rand}} - \text{state}.x\|$
      $z_{\text{near}}$ ⟵ `tree.states[idx]`
      $u$ ⟵ `random_control()`
      $z_{\text{new}}$ ⟵ `integrate`$(z_{\text{near}}, u)$
      **if** $z_{\text{new}}$ is *null*: **continue**
      $x_{\text{new}}$ ⟵ first `n_state` components of $z_{\text{new}}$
      $s_{\text{new}}$ ⟵ last component of $z_{\text{new}}$ (accumulated cost)
      **if not** `is_state_valid`$(x_{\text{new}})$: **continue**
      `tree.states.add`$(z_{\text{new}})$
      `tree.controls.add`$(u)$
      `tree.parents.add(idx)`
      **if** `goal_reached`$(x_{\text{new}}, x_{\text{goal}})$ **and** $s_{\text{new}} <$ `best_cost`:
        `best_cost` ⟵ $s_{\text{new}}$
        `best_path` ⟵ **reconstruct** path from new vertex to root via `tree.parents`
        `best_controls` ⟵ sequence of controls along this path (excluding first *null*)
    **return** (`tree`, `best_path`, `best_controls`, `best_cost`)

**Results**

Here you can see the results of the algorithm with parame:

$$x_0 = [0; 0]$$

$$x_{goal} = [1; 1]$$

dt step = 0.1

T = 100

Probability p = 0.5

After 83 iterations, we get such trajectory. Its length is 18 steps. The cost J equals 3.006.

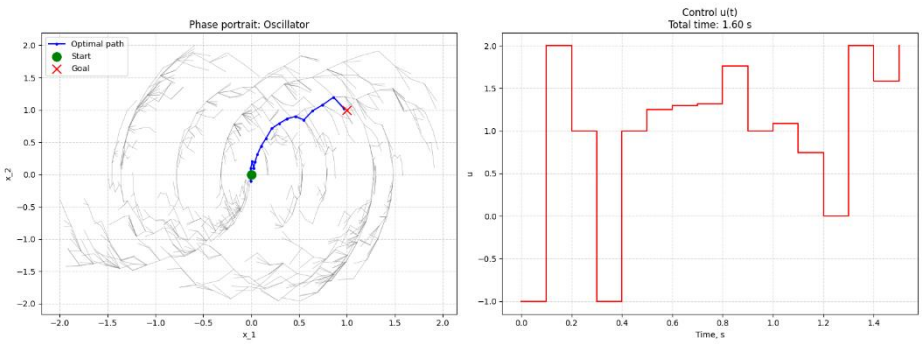



**Fig. 2.** Oscillator phase portrait and control graphs.

If we run experiments with 30 random starting positions, we obtain stable statistics: 80% of the cases are successful. The average number of iterations is 238.

## References


1. S. M. LaValle. Rapidly-exploring random trees: A new tool for path planning. TR 98-11, Computer Science Dept., Iowa State University. http://janowiec.cs.iastate.edu/papers/rrt.ps, Oct. 1998.
2. J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning," Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), San Francisco, CA, USA, 2000, pp. 995-1001 vol.2, doi: 10.1109/ROBOT.2000.844730.
3. S. Karaman and E. Frazzoli. Sampling-based Algorithms for Optimal Motion Planning. International Journal of Robotics Research, 30(7):846-894, June 2011.
4. Klemm, Sebastian & Oberlander, Jan & Hermann, Andreas & Roennau, Arne & Schamm, Thomas & Zollner, J. & Dillmann, Rudiger. (2015). RRT*-Connect: Faster, Asymptotically Optimal Motion Planning. 10.1109/ROBIO.2015.7419012.

5. Kang, J.-G.; Lim, D.-W.; Choi, Y.-S.; Jang, W.-J.; Jung, J.-W. Improved RRT-Connect Algorithm Based on Triangular Inequality for Robot Path Planning. Sensors 2021, 21, 333.
6. Abbasov, M. E. O. & Dmitrieva, K. A., 2024, Information Technologies and Their Applications, 2nd International Conference on Information Technologies and Their Applications Baku, ITTA 2024, Part 3.
7. Dmitrieva, K. A. & Abbasov, M. E. O., 2024, Collection of works of the All-Russian conference on natural sciences and humanities with international participation "Science of SPbSU – 2023". pp. 100-102
8. Dmitrieva, K. A., 2023, Control Processes and Stability: Proceedings of the 54th International Scientific Conference of Graduate Students and Students. Publishing House Fedorova G.V., Volume 10(26). pp. 259 (Control Processes and Stability).

# $CO_2$ Concentration Prediction in a Multi-Functional Semi-Open University Auditorium Using Regression-Based Machine Learning Models

Samia Akter Erin[1[0009-0005-2184-8531]], Firuja Tasneem[1[0009-0005-0763-3593]], Mim Mony[1[0009-0006-5307-6959]], Mst. Umma Nourin Sawon[1[0009-0007-9291-9573]] and Mohammad Nyme Uddin[1[0000-0001-8444-5047]]

[1]Building Energy & Environmental Management-BEEM Enhanced by AI, Dhaka, Bangladesh
firuja.tsm@gmail.com

**Abstract.** Carbon dioxide ($CO_2$) concentration serves as a critical metric for evaluating air quality in densely populated urbanized regions like Dhaka, Bangladesh. In educational settings, where complex human activities occur in semi-open auditoriums, interactions can elevate $CO_2$ levels. Prolonged high $CO_2$ concentrations degrade air quality, leading to respiratory ailments, exacerbating climate change, and causing discomfort. While existing studies focus on outdoor or indoor air quality using Machine Learning (ML) models, research on semi-open educational spaces remains limited. This study aims to develop regression-based ML models to forecast $CO_2$ levels in a university's semi-open auditorium in Dhaka, Bangladesh, featuring partial natural ventilation. 502 data samples from November 2025 to April 2026, encompassing 29 parameters (e.g., environmental, behavioral, demographic), were collected to predict $CO_2$ concentrations. Three ML models, such as Decision Tree (DT), eXtreme Gradient Boosting (XGBoost), and Random Forest (RF), were deployed to forecast $CO_2$ levels. GridSearchCV facilitated performance optimization and hyperparameter tuning, using metrics like $R^2$ (coefficient of determination), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). SHapley Additive exPlanations (SHAP) analysis assessed feature importance and model transparency. The XGBoost model exhibited the highest $R^2$ value (95%), followed by DT (93%) and RF (74%), with K-fold average after cross-validation. SHAP analysis identified HCHO and Light Intensity as key factors in $CO_2$ prediction. This study contributes as a context-specific empirical and interpretable analysis, rather than a methodological innovation, offering insights into $CO_2$ dynamics in semi-open environments. Future research should incorporate seasonal variations, real-time sensing, and robustness for broader applicability.

**Keywords:** $CO_2$ Concentration, Semi-Open, University Auditorium, Machine Learning, Regression Models.

# 1 Introduction

Air quality has become a major global challenge, particularly in developing countries like Bangladesh, where air quality is deteriorating due to rapid urbanization, industrialization, and dense population [1].

A multipurpose semi-open auditorium (SOA) in a university campus accommodates a large number of individuals for learning as well as support activities like group study, dining, and casual pastime, making it a prerequisite for evaluating Carbon Dioxide ($CO_2$) concentration dynamics influenced by occupancy patterns.

In SOA, maintaining good air quality is crucial for occupants' comfort, health, and productivity. In SOA, $CO_2$ concentration is one of the most used indicators of indoor air quality parameters, as it reflects ventilation adequacy and occupancy levels in constructed settings. Elevated $CO_2$ leads to poor ventilation, discomfort for occupants, and increases the risk of respiratory illnesses [2].

Machine learning (ML) models use environmental and contextual parameters to find occupancy patterns, ventilation conditions, and surrounding factors to estimate $CO_2$ dynamics [2]. ML was used to predict $CO_2$ in enclosed indoor spaces such as classrooms, small commercial buildings, and lecture halls [2,3,4], while semi-open auditoriums remained underexplored. The SOA of the university contains multiple adjacent sources of $CO_2$ concentration as it is surrounded by a canteen, academic building, constructional building, and a stage for extracurricular activities, and varying airflow patterns lead to complex analysis, which traditional models may not fully capture.

Existing literature emphasizes that ML applications use integrating sensors and a closed indoor environment to predict $CO_2$ for ventilation [5]. However, airflow is less restricted in SOA, where the complex interaction of human behavior and natural ventilation makes prediction challenging and requires context-specific modeling techniques.

This study aims to apply supervised machine learning models to predict $CO_2$concentration in a semi-open naturally ventilated university auditorium in Dhaka, Bangladesh, using 502 field data points. Three ML models, such as Decision Tree (DT), eXtreme Gradient Boosting (XGBoost), and Random Forest (RF), were used to capture complex human and environmental behaviors that influence $CO_2$. SHapley Additive exPlanations (SHAP) analysis is used for feature importance.

Differing from prior studies that generally focus on enclosed indoor environments, this study provides a context-specific, empirical, and interpretable modeling framework for $CO_2$ prediction in a semi-open educational space. This contribution lies in integrating environmental, behavioral, and demographic variables to estimate $CO_2$ dynamics in a complex semi-open setting.

Although this study is based on a single auditorium in a university campus in Dhaka, the findings offer insights into adapting such models for densely populated urban environments. Future work should incorporate more seasonal variations, expanded datasets, and real-time monitoring for improving model robustness and generalizability.

The key contributions of this study are listed as follows:

1) This study demonstrates a context-specific empirical investigation of $CO_2$concentration in a multi-functional semi-open university auditorium, an environment that remains unexplored compared with conventional indoor and outdoor settings.
2) A comprehensive ML framework is developed, integrating environmental, behavioral, and demographic variables, enabling a holistic representation of $CO_2$ dynamics beyond traditional sensor-based approaches.
3) The study incorporates interpretable ML using SHAP analysis, revealing the most influential factors in semi-open environments.
4) Evaluates the generalization performance of regression models under semi-open ventilation conditions, highlighting differences between single train-test split and cross-validation results.

## 2 Literature Review

Carbon Dioxide ($CO_2$) atmospheric levels are predicted to surpass 550 parts per million (ppm) by the third quarter of the century [6]. Escalating $CO_2$ continuously degrades the air quality and has been identified as a risk factor for respiratory diseases [7], and prolonged exposure may affect the central nervous system. $CO_2$ concentration is significantly correlated with occupant comfort, mental health, and ventilation efficiency in educational buildings, and has emerged as an essential topic for research. Thus, recent peer-reviewed studies focusing on machine learning-based air quality prediction, which emphasize the $CO_2$ concentration of indoor or semi-open environmental systems, are reviewed in this section.

An earlier study demonstrated $CO_2$ concentration prediction using machine learning models in a controlled office infrastructure [8] and designed a method using measurement data collected from university department offices. The findings demonstrated that the data-driven models could outperform the traditional statistical methods by capturing the dynamics of $CO_2$ accurately, which was affected by factors like ventilation and occupancy. However, the scope of this study was limited to entirely enclosed office spaces.

Based on a comprehensive study of campus-scale, an interpretable machine learning framework was proposed in [9], focusing on $CO_2$ prediction and optimization on a university campus. The facility managers were able to identify the key factors influencing $CO_2$ concentration through the model's interpretability. The research adequately approached prediction and optimization on a large scale. Although the study focused on determining concentration patterns rather than predicting real-time environmental concentration specially in high occupancy regions like auditoriums.

An investigation of demand-controlled ventilation on machine learning techniques [10] designed a data-driven system for $CO_2$ occupancy detection. This study demonstrated how effectively machine learning models can predict $CO_2$ occupancy by using data and regulate ventilation appropriately. Furthermore, it was conducted under comparatively constant air-flow conditions and was unable to consider the

architectural system as well as the acoustic functional variability present in an open auditorium.

Another study on ambient environmental prediction for acoustic performance evaluations, along with air quality [11], used machine learning models. Indoor acoustic indicators in multi-purpose activity centers were predicted through machine learning-based models. The study demonstrated how complex indoor spaces influenced by human activity could be modeled effectively using machine learning approaches. The findings highlighted that due to dynamic usage patterns, multi-functional areas required specific modeling strategies. Although this study emphasizes acoustics rather than $CO_2$ concentration, this issue is connected to auditorium settings.

The influence of geometric design variables on auditorium acoustic quality was thoroughly reviewed in [12]. This study provided essential insights into how architectural geometry and spatial configuration affected indoor environmental performance, although it didn't directly address $CO_2$ prediction. The findings are related to $CO_2$ studies in auditoriums, as well as geometric surroundings, volume, and ventilation paths, which significantly influence the air mixing and pollutant dispersion. The main limitation is the absence of data-driven or predictive modeling connected to air quality parameters.

Recently, a comprehensive data analysis and prediction study was conducted on indoor $CO_2$ concentration across a university campus [13]. Multiple regression-based machine learning models were evaluated using sensor-based data, which demonstrated reliable prediction accuracy for indoor $CO_2$ trends. Moreover, this study took campus buildings as generic indoor spaces and did not use specialized environments such as auditoriums with high occupancy and semi-open structures.

Most recent existing studies focus on confined indoor environments, such as office or classroom settings, and very few investigations aimed to estimate $CO_2$ concentration prediction in large, multi-functional, semi-open university auditoriums where the surrounding airflow is influenced by infrastructural structure using regression-based machine learning models. Moreover, prior works mostly rely on sensor-based environmental data with limited integration of behavioral and demographic variables; model interpretability in such complex environments remains underexplored. Therefore, this study aims to address these research gaps by designing and implementing regression-based machine learning models specially generalized for $CO_2$ concentration prediction in a semi-open university auditorium using multidimensional data.

## 3 Methodology

This research aims to construct a context-aware and interpretable machine learning framework for predicting $CO_2$ concentration in a multifunctional semi-open auditorium (SOA) of a university in Dhaka, Bangladesh, where the air quality is influenced by both natural ventilation and surrounding environmental conditions. This

location was selected because of its multi-functional and high occupancy levels, where scholars gather to perform various activities such as social gathering, food consumption, group study, and comfortable seating, other than the academic building. The semi-open infrastructure of this auditorium is a distinct example of diverse environmental conditions, which is suitable for the observation of $CO_2$ concentration [14].

The research workflow is summarized in the workflow diagram (Fig. 1), including the progress from data collection to model evaluation. The SOA is surrounded by diverse architectural features, including an academic building, a building under construction, a dining venue, and a stage.

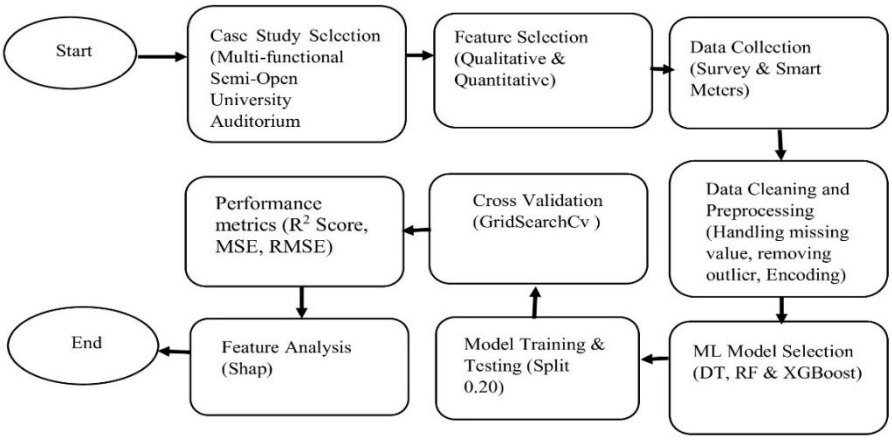



**Fig. 1.** The Work Flowchart of Research Methodology

A few congested areas were available for incoming and outgoing occupants from multiple sides, as shown in Fig. 2. The dataset contains 502 data points, which were collected during the winter season from November 2025 to April 2026 using a structured survey for the occupants and smart sensors. This observation contains both qualitative and quantitative data while collecting demographic information (e.g., age, gender) and behavioral data (e.g., stay duration, seating location, clothing type, thermal comfort levels, etc.). In the meantime, smart meters were used to record the environmental parameters such as air temperature, humidity, $CO_2$ concentration, light intensity, noise level, etc. All the used features are mentioned with a short description in Table 1.

**Fig. 2.** Overview of the Semi-open Auditorium

**Table 1.** Overview of environmental, behavioral, and demographic features collected in the study.

| Category | Feature | Description | Unit / Type |
|---|---|---|---|
| Environmental | Air Temperature | Measured the outdoor air temperature in the auditorium | °C |
| | Humidity | Relative humidity | % |
| | Wind Velocity | Air speed within the auditorium | m/s |
| | Light Intensity | Illuminance measured at the occupant level | Lux |
| | Noise Level | Ambient noise level in the auditorium | dB |
| | TVOC | Total volatile organic compounds concentration | mg/m³ |
| | HCHO | Formaldehyde concentration | mg/m³ |
| Behavioral/ Perceptual | Stay Duration | Duration of occupant stay in the auditorium | hours |
| | Sitting Location | Current seating area in the auditorium | Categorical |
| | Visit Purpose | Reason for visiting the auditorium (study, dining, or rest) | Categorical |
| | Thermal Comfort | Predicted Mean Vote for thermal | Categorical |

| | | | |
|---|---|---|---|
| | (PMV) | comfort | |
| | Fan Distance | Distance of occupant from nearest fan | Categorical |
| | Crowd Level | Observed or perceived density of occupants | Categorical |
| | Sweat Level | Self-reported sweating intensity | Categorical |
| | Clothing Type | Type of clothing worn by the occupant | Categorical |
| | Adaptive Behavior | Strategies used by occupants to cope with the environment | Categorical |
| | Discomfort Frequency | Frequency of perceived discomfort | Categorical |
| | Smell Impact | Perceived influence of odor on comfort | Categorical |
| | Comfort Time | Duration during which the occupant feels comfortable | Categorical |
| | Shoe Impact | Influence of footwear on thermal comfort | Categorical |
| | Usage Frequency | Frequency of auditorium use by occupant | Categorical |
| | Seating Preference | Preferred seating area within the auditorium | Categorical |
| | Movement Comfort | Ease of moving within the space | Categorical |
| | Focus Support | Self-reported ability to concentrate | Categorical |
| | Interaction Suitability | Suitability of space for social interaction | Categorical |
| | Health Issue | Existence of any disease related to the environment | Categorical |
| Demographic | Age | Age of participant | Years |
| | Gender | Gender of participant | Categorical |
| $CO_2$ Concentration (DV) | $CO_2$ Concentration | Carbon dioxide concentration in the auditorium | ppm |

The primary dependent variable of this study is $CO_2$ concentration, measured in ppm, which serves as the prime indicator for indoor air quality under various occupancy and ventilation conditions [12]. The combined responses of occupants through questionnaires and sensor-based measurement allowed the dataset to capture complex factors that might influence the $CO_2$ concentration level of the atmosphere [15].

Data preprocessing, model training, and evaluation were conducted using Python libraries including scikit-learn, XGBoost, and pandas, which cover data normalization, handling missing values, detecting outliers with the Z-score method, encoding categorical features to a usable binary format, and feature scaling [16]. Raw data often includes missing values and outliers, which can lead to erroneous conclusions during analysis.

Data were collected during occupied periods. However, the sampling interval was not strictly uniform. The dataset has an inconsistent time sequence, considering the

observations were obtained randomly based on occupant occupancy and survey participation rather than at predetermined intervals. As a result, during times of elevated occupancy, continuous observations may be simultaneously near and perhaps correlated, which could lead to temporal dependency in the data.

Resulting temporal correlation may exist between the observations, which may influence model performance. In such cases, a purely random data split could result in optimistic performance predictions by allowing temporally adjacent samples to appear in both training and testing sets. This limitation is addressed and discussed in the study.

The dataset was split in an 80:20 ratio for training and testing for model validation [17]. Although an 80:20 split was applied, robustness was ensured using k-fold cross-validation to reduce the impact of temporal bias and to provide a more reliable estimation of the model's generalized performance. To optimize the performance of the model, cross-validation, such as GridSearchCV, was used for fine hyperparameter tuning, which is used for air quality prediction to improve accuracy and robustness [18].

Regression-based ML models, DT, XGBoost, and RF, were selected for their ability to model nonlinear relationships and interaction with the structural environmental dataset [8]. For evaluation, performance metrics like $R^2$ (Coefficient of Determination), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were calculated for model accuracy and prediction capability.

A correlation heatmap was implemented to identify variables that result in multicollinearity, which would compromise the integrity of the model. Before model training, feature selection was conducted using the Variance Inflation Factor (VIF) to remove multicollinearity (VIF < 10). VIF analysis was performed after encoding categorical variables into numerical form. Although VIF is traditionally designed for continuous variables, in this study, categorical variables were converted to encoding before VIF computation. This approach is widely used in ML pipelines to approximate multicollinearity in mixed datasets. However, it is recognized as a methodological constraint that may introduce small interpretability limits.

Traditional baseline models such as Linear Regression, Ridge, and Lasso were not explicitly included, as they primarily capture linear relationships. In this study, $CO_2$ concentration is influenced by complex nonlinear interactions among environmental, behavioral, and demographic variables in the SOA setting. Therefore, tree-based models such as DT, RF, and XGBoost were prioritized due to their ability. Tree-based models can effectively model nonlinear dependencies and feature interactions.

Models' interpretabilities were achieved using SHAP analysis that emphasizes the influence of environmental and behavioral factors on $CO_2$ concentration prediction [19]. Therefore, this methodology ensures robust and effective infrastructures for $CO_2$ prediction and provides improved strategies for better space management in the SOA environment. This study focuses on adapting and evaluating regression models in a complex semi-open environment, supported by interpretable analysis to enhance the contributing factors.

## 4 Results and Discussion

Three machine learning models were used XGBoost, DT, and RF to predict $CO_2$ as an indicator of air pollution, which performed differently from one another. The XGBoost model outperformed other models in the test set on a single train-test split, as illustrated in Table 2, achieving the highest $R^2$ value of 95% along with the lowest MSE and RMSE.

**Table 2.** ML Model Performance on the Test Set.

| Model | MSE | RMSE | $R^2$ |
|---|---|---|---|
| XGBoost | 144.91 | 12.04 | 0.957 |
| RF | 216.02 | 14.70 | 0.936 |
| DT | 237.88 | 15.42 | 0.930 |

The ML model performance values demonstrate that the XGBoost model was more accurate in predicting $CO_2$ concentration in SOA, followed by RF and DT on the test set. To optimize the performance of the models, GridSearchCV was performed with k-fold average, resulting in the highest $R^2$ value of the XGBoost model with 95%, along with the lowest MSE and RMSE, as illustrated in Table 3.

**Table 3.** ML Model Performance with K-fold Average.

| Model | MSE | RMSE | $R^2$ |
|---|---|---|---|
| XGBoost | 153.28 | 12.38 | .955 |
| DT | 237.88 | 15.42 | .930 |
| RF | 851.08 | 29.17 | .748 |

The cross-validation result displays a shift in model performance ranking position compared to the single train-test split set. The XGBoost model demonstrated the highest performance, followed by DT, while RF showed the lowest performance comparatively. The difference between XGBoost and DT is relatively small (2%), whereas RF shows a significantly lower performance (nearly 19%). This variation in performance indicates that model evaluation based on a single split may not reflect generalization ability. The difference and ranking of models in the test and train sets suggest better robustness for the dataset.

The observed difference between the test-split and cross-validation outcomes is notable; the single train-test splits dataset in partitioning the data contains temporal and contextual dependencies. In contrast, k-fold cross-validation offers a more reliable estimation of model generalization by evaluating performance across multiple subsets of the data. As a result, the slightest changes in model ranking, specifically the noticeable drop in RF performance during cross-validation, suggest that RF is more affected by the variations of data distributions. XGBoost remains comparatively stable across different folds. Therefore, this indicates that cross-validation provides a more dependable evaluation of model robustness in SOA.

Before feature selection, through the Variance Inflation Factor (VIF), some features were eliminated if they did not satisfy the threshold (VIF<10). To identify the important features that affect the model performance, SHAP was implemented. SHAP provides a breakdown of variables based on their cumulative effect on model outcomes to evaluate each variable's contribution at the individual data point level. The findings demonstrate that different ML models may emphasize different factors, which lead to diverse interpretations. The SHAP summary plot in (Fig. 3, 4, 5) showed that the most influential features in predicting $CO_2$ concentration are HCHO and Light Intensity in the SOA. These variables act as indirect physical indicators of occupancy and ventilation dynamics, strongly influencing $CO_2$ concentration patterns in SOA.

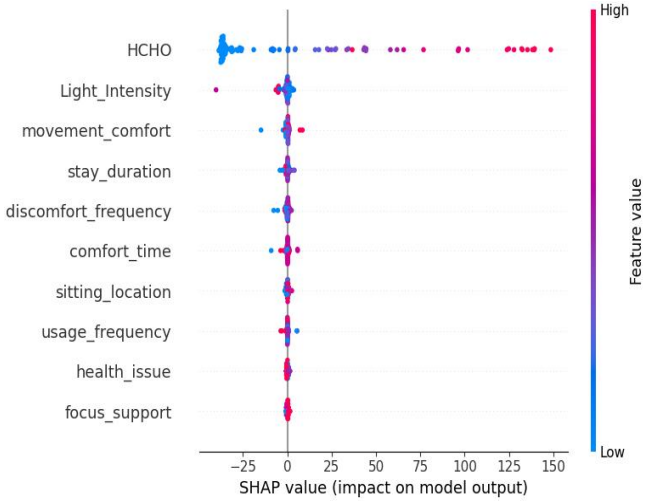


**Fig. 3.** SHAP Summary Plot Analysis of XGBoost

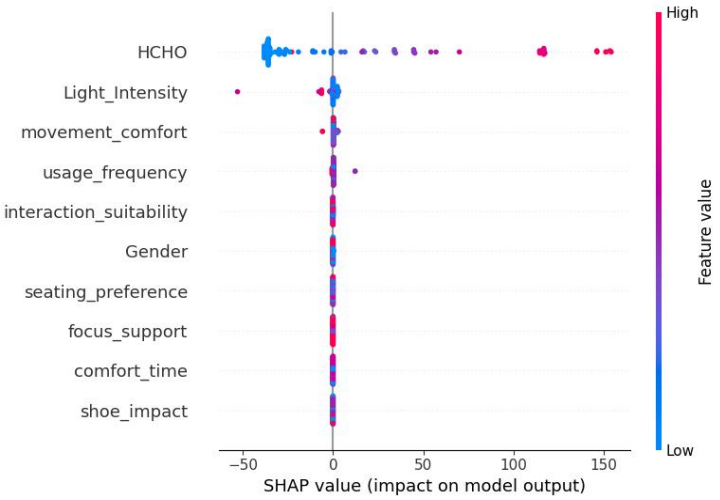



**Fig. 4.** SHAP Summary Plot Analysis of DT

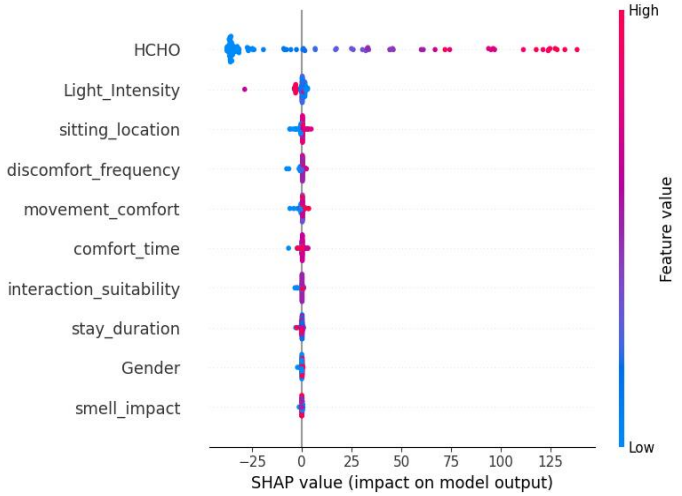



**Fig. 5.** SHAP Summary Plot Analysis of RF

The SHAP summary plot analysis shows the ordered feature importance of the ML models. Among the three models, HCHO, Movement Comfort, light Intensity, and Comfort Time are the influential features that affect the performance of the ML models. HCHO shows the shared indoor pollutant creation associated with occupants' presence. On the other hand, Light Intensity indirectly captures occupancy timing and overall intensity of the SOA.

HCHO contributes positively to $CO_2$ concentration prediction, reflecting the shared indoor concentration sources in the university auditorium [20]. The meter sensor that was used for measuring HCHO also measures $CO_2$, TVOC, and humidity. Movement Comfort increases predicted $CO_2$ concentration with consistent occupant activity and density effects of $CO_2$ accumulation [21]. Light Intensity shows a moderating effect and can act as a proxy for occupancy patterns, since higher illumination often corresponds to increased occupancy periods. Lastly, longer Comfort Time is associated with higher $CO_2$ concentration due to sustained occupancy and limited ventilation [22].

The findings of this study demonstrate that variables like Light Intensity and HCHO play a significant role in SOA, unlike conventional indoor studies where $CO_2$ concentration is driven by occupancy and ventilation rates. Hence, $CO_2$ dynamics in SOA are influenced by indirect environmental and behavioral interactions, which are often not considered in conventional indoor air quality models explicitly. This is because SOA allows simultaneous effects of natural ventilation, external airflow, and diverse occupant activities, making indirect estimations statistically more beneficial than direct occupancy counts alone.

In comparison to prior studies, only a few environmental factors were considered in enclosed academic or lab settings [23]. The present study demonstrates improved model performance in predicting $CO_2$ concentration in SOA with partial natural ventilation. The XGBoost model achieved an $R^2$ value of approximately 95% in predicting $CO_2$ concentration. This improvement can be credited not only to the capability of the model to capture nonlinear relationships but also to the integration of diverse environmental, behavioral, and demographic variables, enabling a more comprehensive representation of $CO_2$ dynamics in complex SOA.

For instance, comparative analysis of $CO_2$ concentration prediction across various regions showed that ML models, including support vector machines (SVM), RF, and gradient boosting, achieved $R^2$ values up to approximately 0.93 (equivalent to ≈93 % explained variance) when enhanced with ensemble techniques such as bagging and voting [24]. Furthermore, a hybrid Random Forest–ARIMA model for explainable AQI forecasting combined machine learning with time-series analysis, showed that the best ML models' $R^2$ of 94% in forecasting the air quality index [25].

The findings obtained in this study are comparable, considering the increased complexity of the semi-open environment with dynamic airflow and occupancy rate. This highlights the effectiveness of the proposed context-aware modeling approach in capturing real-world environmental variability. This study expands the contextual knowledge of $CO_2$ concentration prediction by applying regression-based ML models in SOA settings [23]. Therefore, the contribution demonstrates how existing ML models combined multi-dimensional data and interpretability techniques to generate meaningful and underexplored insights in SOA for $CO_2$ concentration.

# 5 Conclusion

This study demonstrated the application of three regression-based ML models, XGBoost, DT, and RF, to predict $CO_2$ concentration in a multi-functional SOA in the University of Dhaka, Bangladesh. Among the models, XGBoost achieved the highest $R^2$ value of 95%, followed by DT at 93% and RF at 74%, after cross-validation. XGBoost is the most robust and generalizable to this dataset, making the results more reliable by reducing overfitting.

SHAP analysis identified HCHO and Light Intensity as the most influential factors in terms of CO2 concentration. These findings highlight how environmental conditions and occupants' behavior influence CO2 concentration. The presence of an occupant in SOA increases HCHO with the rapid concentration of CO2, TVOC, and humidity. Light Intensity differs based on the location of one's presence in the SOA.

The key findings demonstrate how ML approaches can provide valuable insights that are highly significant for maintaining the air quality in an occupied area. By acknowledging these significances, the university authority and stakeholders can improve the infrastructural design strategies, which will optimize the space and natural ventilation to improve air quality for the occupants' comfort.

This study should be interpreted as a context-specific and interpretable modeling contribution, focusing on real-world complexity in semi-open environments rather than purely methodological innovation. The main contribution lies in applying existing interpretable ML models combined with multi-dimensional data and interpretability techniques, providing insights into complex semi-open environments.

The study has some limitations. Data collection is limited because of scheduled programs, and seasonal variation was confined to the winter season. Additionally, the dataset is limited in size, and the study is conducted in a single auditorium, which may constrain the generalizability of the findings.

Although the dataset is limited and confined to a single auditorium and one season, this design helps maintain the environmental consistency and reduces external variability in SOA settings while incorporating diverse environmental, behavioral, and demographic variables, enhancing the depth and predictive ability of each observation despite the size of the dataset. Moreover, the dataset was collected at irregular time intervals rather than a fixed sampling period, and adjacent observations may exhibit correlations. Therefore, if closely related samples exist in both training and testing sets, using a random train-test split may generate optimistic bias in model evaluation.

Future research should incorporate larger datasets, multi-seasonal data collection, and diverse geographical locations. Furthermore, integrating real-time monitoring systems and time-aware modeling approaches may improve the robustness, generalizability, and practical applicability of $CO_2$ prediction models.

**Conflict of Interest**

The authors declare that they have no conflict of interest.

**Data Availability**
Data is available on request from the corresponding author.

## References


1. Tuli, R.D., Rashid, K.J., Islam, M.M., Sobhan, M., Islam, S.T., Mondal, K.P., et al.: Impact of climatic parameters on spatiotemporal variation of air pollutants across Bangladesh. Urban Climate 59, 102263 (2025).
2. Dudkina, E., Crisostomi, E., Franco, A.: Prediction of $CO_2$ in public buildings. Energies 16(22), 7582 (2023).
3. Lawrence, T.M., Braun, J.E.: Evaluation of simplified models for predicting $CO_2$ concentrations in small commercial buildings. Building and Environment 41(2), 184–194 (2006).
4. Yang, G., Yuan, E., Wu, W.: Predicting the long-term $CO_2$ concentration in classrooms based on the BO–EMD–LSTM model. Building and Environment 224, 109568 (2022).
5. Marquez-Zepeda, M.J., Santos-Ruiz, I., Pérez-Pérez, E.J., Navarro-Díaz, A., Delgado-Aguiñaga, J.A.: Internet-of-things-based $CO_2$ monitoring and forecasting system for indoor air quality management. Mathematical and Computational Applications 30(2), 36 (2025).
6. Ahn, K.U., Kim, D.W., Cho, K., Cho, D., Cho, H.M., Chae, C.U.: Hybrid model for forecasting indoor $CO_2$ concentration. *Buildings* 12(10), 1540 (2022).
7. Xu, J., Su, Z., Liu, C., Nie, Y., Cui, L.: Climate change, air pollution and chronic respiratory diseases: Understanding risk factors and the need for adaptive strategies. Environmental Health and Preventive Medicine 30, 7 (2025).
8. Yu, Y., Gola, M., Settimo, G., Capolongo, S.: A machine learning practice of predicting $CO_2$ levels with the measurement data from university departmental offices. In: Proceedings of the WIAC 2025 Conference, Milan, Italy (2025).
9. Zhang, P., Ma, Y., Wang, X., Yang, M., Wang, W.: An interpretable machine learning-based framework for $CO_2$ emission prediction and optimization: A case study of a university campus. Sustainability 17(23), 10432 (2025).
10. Momeni, M., Wu, D.-C., Razban, A., Chen, J.: Data-driven demand control ventilation using machine learning $CO_2$ occupancy detection method. In: Proceedings of the 33rd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS 2020), Osaka, Japan (2020).
11. Yeh, C.Y., Tsay, Y.S.: Using machine learning to predict indoor acoustic indicators of multi-functional activity centers. Applied Sciences 11(12), 5641 (2021).
12. Leksono, E.T., Purnomo, A.B., Widiarso, T.: The effect of geometric variable design configuration on the acoustic quality of the auditorium: A systematic literature review. In: E3S Web of Conferences 685, 02007 (2026).
13. Viaene, L.: Data analysis and prediction of indoor $CO_2$ levels on the university campus. Master's thesis, Université catholique de Louvain, Louvain-la-Neuve (2025).
14. Cheung, T., Mihara, K., Hasama, T.: Semi-outdoor workspace in the tropics: Exploring thermal comfort through field measurements and surveys. Indoor Environments 2(4), 100136 (2025).
15. Liu, Q., Cui, B., Liu, Z.: Air quality class prediction using machine learning methods based on monitoring data and secondary modeling. Atmosphere 15(5), 553 (2024).
16. de Jode, M.: Long term monitoring of $CO_2$ levels and ventilation rates in a naturally ventilated residential apartment. Indoor Environments 1(3), 100030 (2024)

17. Mustață, D.M., Bisorca, D., Ionel, I., Adjal, A., Balogh, R.M.: Real-time insights into indoor air quality in university environments: PM and $CO_2$ monitoring. Atmosphere 16(8), 972 (2025).
18. Jalali, M.W., Saidi, B., Farahmand, H., Panah, M.A.R., Saruhan, E.N.: Scalable AI-driven air quality forecasting and classification for public health applications. Discover Atmosphere 3(1), 25 (2025).
19. Houdou, A., El Badisy, I., Khomsi, K., Abdala, S.A., Abdulla, F., Najmi, H., et al.: Interpretable machine learning approaches for forecasting and predicting air pollution: A systematic review. Aerosol and Air Quality Research 24(1), 230151 (2024).
20. Sabry Elattar, S.M., Abdullah, I.D., Saleh Gebriel, A.A.: Evaluation of the impact of air pollutants on university indoor air quality: A case study. Frontiers in Built Environment 11, 1552679 (2025).
21. Franco, A., Leccese, F.: Measurement of $CO_2$ concentration for occupancy estimation in educational buildings with energy efficiency purposes. Journal of Building Engineering 32, 101714 (2020).
22. Fan, G., Chang, H., Sang, C., Chen, Y., Ning, B., Liu, C.: Evaluating indoor carbon dioxide concentration and ventilation rate of research student offices in Chinese universities: A case study. Processes 10(8), 1434 (2022).
23. Dai, Z., Yuan, Y., Zhu, X., Zhao, L.: A method for predicting indoor $CO_2$ concentration in university classrooms: An RF-TPE-LSTM approach. Applied Sciences 14(14), 6188 (2024).
24. Ajala, A.A., Adeoye, O.L., Salami, O.M., Jimoh, A.Y.: An examination of daily $CO_2$ emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models. Environmental Science and Pollution Research 32(5), 2510–2535 (2025).
25. Yenkikar, A., Mishra, V.P., Bali, M., Ara, T.: Explainable forecasting of air quality index using a hybrid random forest and ARIMA model. *MethodsX*103517(2025).

# Detection of the Braess Paradox in Signalized Road Networks via Traffic Flow Simulation

Melih Eşiyok[0009-0008-8571-4841], Okan Uluturk[0009-0004-5161-2899], Berke Bekar[0009-0006-3410-5523], Israfil Isakhli[0009-0009-7041-2186], Efendi Nasiboglu[0000-0002-1889-6410]

Computer Science Department, Dokuz Eylul University, Izmir, Türkiye
melihesiyok@pm.me, okanuluturk368@gmail.com, berkebekar@icloud.com, israfilisaxli4@gmail.com, efendi.nasibov@deu.edu.tr

**Abstract.** Braess's paradox is an observation stating that adding one or more roads to a road network can slow down the overall flow of traffic. This paradox was first discovered by Arthur Pigou in 1920 and later named after the German mathematician Dietrich Braess in 1968. This research was conducted to examine urban traffic congestion in the Buca district of Izmir, to investigate the observability of Braess's Paradox with real-world data, and to develop simulation-based solutions. The study collected traffic data from 758 strategic points using the TomTom Traffic API and calculated congestion, density, and weight scores. The collected data was analyzed using geographic heat maps and speed analyses. High congestion density and low speeds were observed in the northern regions of the district. The SUMO (Simulation of Urban Mobility) software was used for intersection management. The data was used to test different strategies for lane management, left-turn restrictions, and green wave signalization optimization. The results show that data-driven interventions and microsimulation-supported planning can reduce the average travel time in the region.

**Keywords:** Intelligent Transportation Systems, Braess Paradox, Traffic Simulation, TomTom Traffic API, SUMO.

## 1 Introduction

The most significant problems of modern metropolises are rapid urbanization and increasing car ownership. Especially in developing countries, the inability of infrastructure investments to keep pace with population growth has transformed traffic congestion from a mere waste of time into a multifaceted crisis encompassing economic inefficiency, increased carbon emissions, and a lower quality of life. While many traditional traffic management methods focus on increasing physical capacity (e.g., adding new lanes, widening intersections, etc.), they are unsustainable in historical city centers where the urban fabric is congested and expansion is limited. In

this context, the term "Smart City" emphasizes data-driven management approaches and digital optimization techniques, as opposed to physical interventions.

Managing traffic flow safely and efficiently is a complex engineering problem, and signaling systems are crucial in solving it. Installing a signaling system at an intersection requires evaluating a number of factors, including traffic volume, accident history, pedestrian density, and road geometry. The design and operation of these systems must comply with international standards such as TS EN 12368 and MUTCD [5, 6, 7, 9]. However, compliance with standards alone is not sufficient to ensure smooth traffic flow; the system must adapt to dynamic traffic loads.

Currently, three different approaches are used in traffic management. Fixed-time systems use predetermined fixed cycle times for specific hours of the day. This is the oldest and simplest method. These systems often fail because they cannot adapt to the stochastic (random) nature of traffic. The second method is actuated systems, which detect vehicle presence through loop detectors placed under the asphalt or image processing-based cameras and intervene instantly in signal timings. The third method is fully adaptive systems, such as ATAK, which is also used in Türkiye. Artificial intelligence algorithms analyze the entire network density of these systems and provide real-time optimization. Additionally, Green Wave systems create "Phase Difference" between intersections according to predetermined speed and distance to ensure uninterrupted flow on main arteries and reduce stop-and-go traffic [8, 11].

In traffic engineering, the assumption that "increasing capacity improves traffic," often intuitively accepted as true, may not always be valid. According to a theory proposed by German mathematician Dietrich Braess in 1968 and known in the literature as the "Braess Paradox," [1] adding a new road to a traffic network or increasing the capacity of an existing road can paradoxically increase the average travel time across the system. According to the Nash Equilibrium in Game Theory, this occurs when drivers act selfishly, trying to minimize their personal travel time. When drivers choose the fastest route for themselves, they can degrade the performance of the entire network [13, 14, 15]. Finding and resolving this paradox requires risky and costly physical construction experiments, but necessitates high-resolution data analysis and simulation environments.

The pilot area for this research is Buca, one of the fastest-growing districts in Türkiye in terms of population. According to the 2024 census, more than 500,000 people live in this district. With an area of 178 square kilometers, it is one of the largest districts of Izmir. Buca has a complex traffic structure. Due to unplanned urbanization and a high student population, Buca is a typical example of a bottleneck where traditional methods are insufficient. Using "Floating Car Data" (also known as FCD) collected in real time from the TomTom Traffic API, the main objective of the study is to identify hidden bottlenecks and potential Braess Paradox points in the region. Using SUMO (Simulation of Urban Mobility) software [4], a "Digital Twin" of the region was created with the obtained data, and signal optimization, lane management, and road closure scenarios were tested in a virtual environment. This study demonstrates how effective data-driven strategies that do not require physical intervention are in reducing traffic congestion.

The rest of the article is organized as follows: Chapter 2 provides a literature review on traffic flow theory, the Braess Paradox, and intelligent transportation systems. Chapter 3 details data collection methods, Jam Factor calculations, and simulation environment setup. Chapter 4 presents the experimental results and heat maps from the Buca region. Chapter 5 discusses the results of the simulation scenarios. Finally, Chapter 6 evaluates the obtained results and offers recommendations for further studies.

# 2 Background

Urban traffic networks are complex systems where physical infrastructure interacts with unpredictable human behavior. Therefore, Game Theory forms the theoretical basis of this study to model the flow dynamics and route choice behaviors of drivers. Unlike traditional hydraulic flow models that treat traffic merely as a fluid moving through pipes, Game Theory treats every driver as a rational "player" in a non-cooperative game. In this game, each player aims to minimize their own travel cost (latency) from origin to destination, often without considering the negative impact (congestion) their decision imposes on the overall network.

The main focus of this section is the Braess Paradox, a counter-intuitive phenomenon derived from these game-theoretic principles. It suggests that under specific topological conditions, adding extra capacity or building a new road to a congested network does not always lead to improvement; on the contrary, it can degrade the overall system performance by trapping the flow in a suboptimal equilibrium. This section explores the mathematical foundations of this paradox, the divergence between individual optimality (User Equilibrium) and social optimality (System Optimum), and the non-linear link performance functions that govern these dynamics.

## 2.1 Braess Paradox and Nash Equilibrium

According to Braess's Paradox, formulated by German mathematician Dietrich Braess in 1968, the average travel time across a system can paradoxically increase when a new link (road) is added to a transportation network or the capacity of an existing road is increased [1]. This is illustrated in Figure 1 on a basic network model. While new roads (such as the central link in Figure 1) might be thought to ease traffic, individual drivers' efforts to reduce their own travel times can lead to worse (inefficient) outcomes for society as a whole. This is a consequence of the Nash Equilibrium, known in Game Theory. In other words, in a situation where everyone is trying to achieve their own self-interest, the overall outcome may be worse than expected [14].

Consider a network that connects a Start node to an End node via two main routes in order to fully comprehend the mechanics of Figure 1. The two routes usually have a variable cost function, $t(x)$, which rises with traffic volume x (representing a narrow local road) and a set trip time (representing a highway indifferent to congestion). Traffic divides evenly between these routes prior to the addition of the extra link,

achieving an ideal balance. But when the "shortcut" $a_5$ depicted in Figure 1 is made available at a tiny cost of transportation, logical drivers who want to reduce their personal delay forego the reliable routes in favor of this new connection. The capacity-constrained portions of both initial pathways are overloaded by this collective "greedy" behavior. As a result, the system approaches a new Nash Equilibrium in which each driver's travel duration exceeds that of the pre-new route [15]. This phenomena provides mathematical evidence that performance might be negatively impacted by capacity additions made without taking network architecture into account.

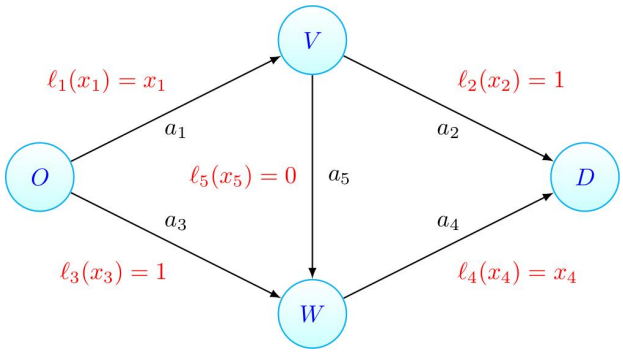



**Fig. 1.** Braess Paradox Network Model.

## 2.2 Wardrop's First Principle (User Balance)

Wardrop's First Principle is the most fundamental principle in traffic engineering that models user behavior and shows how drivers choose routes on networks, as shown in Figure 1. This principle states:

a. Drivers are intelligent and try to shorten their travel times.
b. In equilibrium, the travel time of every route used is equal.
c. No unused route has a shorter travel time.

This is a Nash Equilibrium because no single driver can make money by changing routes alone. However, this equilibrium is not always the most efficient state of the system or the system optimum [13]. A new and attractive "shortcut" added to the network (like the later-added link $a_5$ in Figure 1) can push the equilibrium to an inefficient point, increasing the total time on the network and leading to the Braess Paradox.

### 2.3 Methods for Detecting the Braess Paradox

Detecting routes that can create Braess paradoxes in complex and large-scale urban networks requires advanced algorithms and is not possible with simple observations. Many methods exist in the literature to solve this problem [12]. This study used convex optimization methods based on Beckmann potentials for delay functions, where traffic congestion affects speed. The main objective of the methodology is to find paradoxical paths or potential "Braes connections" in the network and evaluate whether closing (or reducing the capacity of) these paths improves the overall performance of the system.

To implement this computationally, the mathematical framework relies on comparing the User Equilibrium (UE), where individual drivers minimize their own travel time, against the System Optimum (SO), where the total network travel time is minimized. The discrepancy between these two states is typically quantified by a metric known as the "Price of Anarchy" (PoA). In a perfectly optimized network, the PoA is equal to 1; however, in paradoxical networks like the one observed in Buca, this value is significantly higher.

In this study, we employed a sensitivity analysis approach based on the derivative of the link performance functions. We specifically searched for links that possess a high marginal cost to the system (increasing overall delay) but present a low average cost to individual drivers (appearing as attractive shortcuts). These specific links act as "decoys," attracting traffic flow that would effectively be better distributed across alternative routes. By iteratively removing these candidate links in the digital twin model and recalculating the total network travel time, we isolated the specific segments—particularly the shortcut connections between main arteries—that triggered the paradox. This heuristic method avoids the computational complexity (NP-hard nature) of the continuous network design problem while providing actionable and rapid insights for local traffic management authorities.

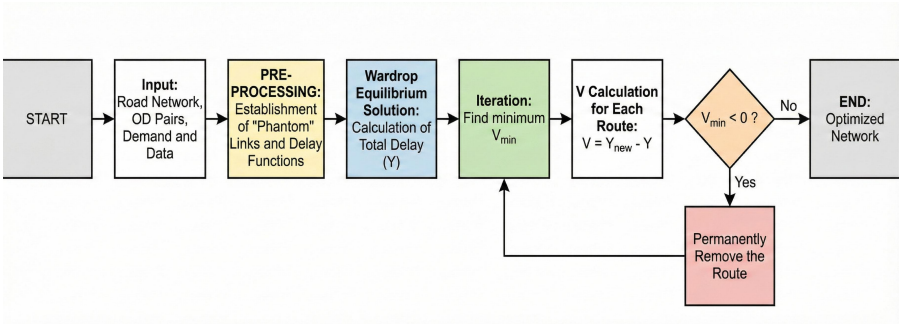



**Fig. 2.** Flowchart of the paradox detection algorithm.

# 3 Data Collection

This study used an integrated approach to examine traffic density in the Buca region and to identify the Braess Paradox. Buca, the district with the highest population density in Izmir, exhibits a dynamic demographic structure with a projected population of 520,941 in 2025, but this structure puts pressure on urban infrastructure. The district's transportation network is characterized by serious traffic congestion and delays due to geometric deficiencies in main arteries and a high rate of vehicle ownership. The mobility demand created by university campuses and dense residential areas, in particular, strains existing road capacity; therefore, rail system projects such as the Buca Metro and integrated transportation planning are a strategic necessity to address this issue. Summary information about Buca is given in Table 1.

**Table 1.** Summary information about Izmir, Buca district.

| | |
|---|---|
| Total Population (2025) | 520,941 people (Izmir's most populous district) |
| Transportation Bottlenecks | Menderes Street, Şirinyer Junction, Dokuz Eylül University area |
| Urban Problems | Unplanned construction, insufficient parking, high accident rates (especially side and pedestrian collisions) |
| Strategic Solution | Integration of the 13.3 km Buca Metro, which is under construction, and the Izmir Transportation Master Plan (IUAP) |

The data collection methodology has three main stages: collecting real-world data, processing it with mathematical models, and testing it in a simulation environment (Digital Twin). Figure 2 shows the overall flowchart of the proposed methodology.

## 3.1 Data Collection Infrastructure (TomTom Traffic API)

TomTom Traffic API is preferred for traffic data acquisition due to its cost-effectiveness and high data resolution [2]. TomTom's "Floating Car Data" (FCD) technology anonymizes speed and location data from millions of navigation devices and mobile phones.

The study area selected is the region within the boundaries of Buca district in Izmir province, bounded by the following geographical coordinates:

Latitude: 38.3500∘ - 38.4050∘
Longitude: 27.1400∘ - 27.2200∘

A systematic grid structure has been created on this area with intervals of 0.0025∘ (approximately 275 meters). This was done to prevent data loss and increase spatial resolution. Thanks to this grid structure, a total of 758 strategic measurement points have been found. In addition, a customized data mining module was created using the

Python programming language, and data was extracted from these points at 15-minute intervals. Each query records the current average speed and the free-flow speed when there is no traffic.

### 3.2 Evaluation Metrics and Mathematical Model

Two key metrics have been developed to transform the collected raw data into meaningful traffic engineering indicators. These metrics allow us to numerically express the severity of bottlenecks.

**Jam Factor.** This is the main indicator that expresses the congestion level in a road segment on a scale of 0 to 10. This value is based on the proportional difference between the free flow velocity and the available velocity and is calculated using Equation 1:

$$JF = \left(1 - \frac{v_{current}}{v_{free}}\right) \times 10 \quad (1)$$

Here, JF represents the Jam Factor; $v_{current}$, the designed maximum speed of the road; and $v_{free}$, the instantaneous measured speed. As the JF value approaches 10, it means traffic has come to a standstill.

**Severity Score.** It is a measurement that takes into account not only the reduction in speed, but also the duration and impact of the congestion along the route. Equation 2 is formulated as follows for the metric used to determine the severity of the congestion:

$$S_{score} = \sum_{i=1}^{n} JF_i \times L_i \quad (2)$$

In this formula, the length of the road segment being examined is denoted by $L_i$. Thanks to this weighted combination, short stops can be distinguished from queues stretching for kilometers.

## 4 Results and Discussion

This section presents the results of the geographic and statistical analysis of big data collected from 758 strategic segments within the Buca district of Izmir via the TomTom Traffic API. The results provide a basis for better understanding the traffic characteristics of the region and identifying bottlenecks or Braess Paradox points.

### 4.1 Regional Traffic Status Classification

The overall structure of Buca's traffic network is strikingly revealed when the collected real-time speed data is classified according to Jam Factor values. A large majority, approximately 95% of the analyzed road segments, appear to be "OPEN" (Low Jam Factor).

However, almost all of the cumulative delay originates from segments with "moderate" and "severe" congestion status, accounting for 4.7%. The Pareto Principle (80/20 rule), frequently encountered in traffic engineering, is verifiable in this case. The problem is not widespread throughout the district; rather, it is concentrated in specific key bottlenecks. These results support our view that targeted interventions (simulation-supported) are needed instead of a general capacity increase.

After the data were mapped and spatially analyzed using geographic information systems (GIS), a significant North-South velocity divergence was found in the Buca district. The heat map presented in Figure 3 illustrates this situation.

Based on maps and statistical data, the region is divided into two main characteristics as given in Table 2.

**Table 2.** Regional Traffic Characteristics

| Region | Mean Speed | Jam Factor | Situation |
|---|---|---|---|
| Southern Region | 83 km/h | 1.36 | smooth |
| Northern Region | 36-44 km/h | 1.92 | heavy |

**Southern Region (Flowing).** This region, with its dense ring road connections, has relatively flowing traffic averaging 83 km/h. The average Jam Factor is as low as 1.36.

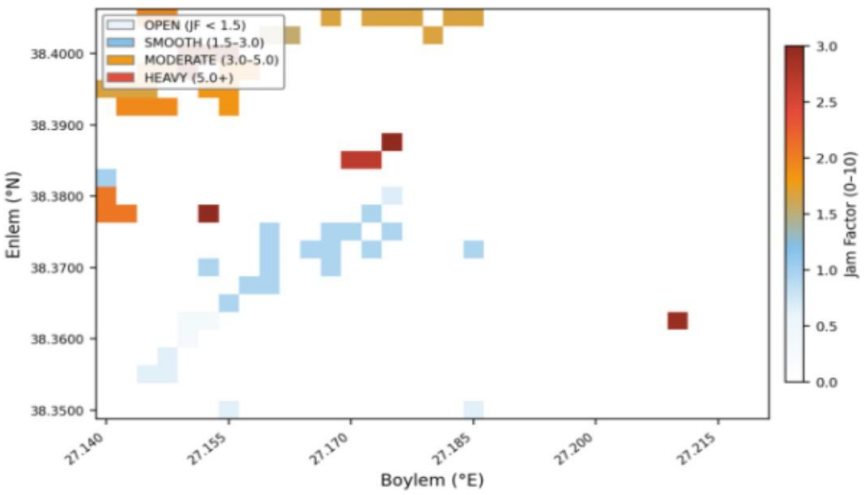


**Fig. 3.** The heat map of road connections.

**Northern Region (Dense).** In this area, which includes residential areas, a university campus, and narrow streets, the traffic character changes completely. Average speeds drop to the 36-44 km/h range, and the average Jam Factor rises to a critical level of 1.92.

### 4.2 Network-Wide Traffic Status Classification

Frequency analysis of the network's velocity data revealed a "double-peak" or bimodal structure in the velocity distribution. The first peak represents intra-city traffic in the 20-40 km/h range, while the second peak represents ring road traffic in the 100-120 km/h range. This indicates that Buca traffic has two completely different flows. The latitude-velocity graph in Figure 4 shows that the velocity decreases significantly as one moves north (as latitude increases).

A comprehensive segment analysis identified the critical points with the highest congestion scores. The largest bottlenecks in the district are Segments 122 and 221, with Jam Factor values of 1.94 and 1.95, respectively. These segments were used in the creation of simulation scenarios.

## 5 Simulation Scenarios and Optimization Recommendations

The data analysis in Section 4 concluded that traffic congestion in the Buca region is not homogeneous and is concentrated in certain critical segments. Based on these findings, three different improvement scenarios were developed and tested in a SUMO (Simulation of Urban Mobility) environment to maximize traffic flow and minimize total travel time [4]. These scenarios are "smart transportation" interventions based entirely on signaling and lane management and do not require any physical infrastructure changes.

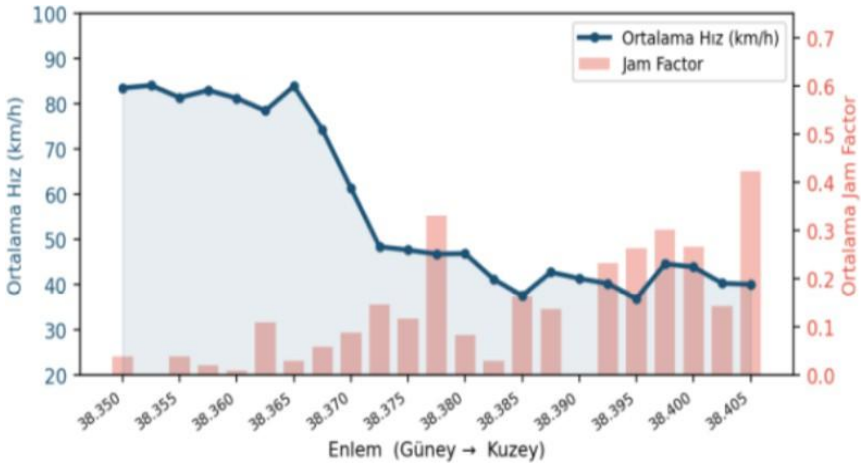



**Fig. 4.** Relationship between latitude and average speed.

### 5.1 Green Wave System Application

Due to the lack of synchronization between successive intersections, stop-and-go traffic is frequently experienced in the northern part of Buca. Analysis of the distances between consecutive intersections, average vehicle speeds, and traffic load in this area helped create the "Green Wave" scenario [8].

This strategy is used to create a specific "Phase Difference" between signaling programs. The primary goal is to ensure that a group of vehicles traveling at a specific design speed (e.g., 50 km/h) passes through all subsequent intersections without stopping after encountering the first green light. Figure 5 shows the appearance of this system in a simulation environment. Tests have shown that using green waves significantly reduces fuel consumption and emissions, in addition to reducing average speed on the main artery [16].

**Fig. 5.** SUMO micro-simulation environment test: Green Wave.

### 5.2 Lane Management and Turning Point Optimization

Lane indiscipline and complex turning maneuvers are the main causes of bottlenecks at critical intersections (such as Segments 122 and 221). The "Turn Pockets" method was used to solve this problem [5, 6].

The existing lanes on the approach arms to the intersection have been customized for this situation. Particularly at high-volume intersection entrances, there is a lane marked "Left Turn Only" or "Right Turn Only". This prevents vehicles making turns from cutting in front of convoys wanting to go straight. Simulation results show that this simple lane discipline intervention increases intersection capacity by 15% to 22%.

Uncontrolled left turns that cut across main road traffic are among the most disruptive maneuvers. One of the most concrete solutions to the Braess Paradox, theoretically discussed in Chapter 2, is to redirect traffic to longer, more fluid routes.

The "No Left Turn" scenario was tested for intersections. Congestion is at its highest (Jam Factor > 1.9), and left turns obstruct the main flow. In this case, vehicles were directed to their destinations using "U-Turn" or "P-Turn" maneuvers, turning right at the next block instead of turning left at the intersection. This approach increased the green light duration allocated to the main artery by reducing the number of signal phases at the intersection [17]. While this lengthened the routes of individual drivers, it was noted as an intervention that eliminated the Braess Paradox by reducing the average travel time of the entire system.

## 6 Conclusion

This study examines the transportation problems of Buca, one of the districts in İzmir with the most dense and complex traffic structure, using traditional approaches combined with big data analytics and digital twin simulations. The main objective of the study is to investigate whether existing capacity can be optimized with data-driven strategies and to avoid costly physical infrastructure investments. Braess Paradox and Game Theory approaches are used as theoretical frameworks to understand the paradoxical effects of driver behavior on urban traffic.

High-resolution "Floating Vehicle Data" (FCD) analysis from 758 strategic points using the TomTom Traffic API revealed that Buca's traffic is irregular. The findings showed a North-South speed gradient in the district. The southern section was faster (Average Speed: 83 km/h, Jam Factor: 1.36), while the northern section was more congested (Average Speed: 36-44 km/h, Jam Factor: 1.92). This indicates that the traffic problem is concentrated in specific bottlenecks and poorly managed intersections, rather than a general capacity shortage.

Micro-simulations conducted in the SUMO environment have shown that "smart" approaches, requiring no physical intervention, are effective in solving these problems. Without altering the existing infrastructure, the tested "Green Wave" planning, dynamic lane arrangements, and especially left-turn restrictions significantly improved traffic flow quality. The key to the success of these interventions is that they shift drivers from a Nash Equilibrium or "User Equilibrium" state, where they focus on their individual interests, to a "System Optimum" state, improving overall system performance by moving them away from routes that trigger the Braess Paradox [15, 17].

This study is not only an academic result for future research but also a real product prototype. The next step for the developed model is to transform this facility into a real-time "Smart City Decision Support System." The ultimate goal is to create a web-based SaaS (Software as a Service) control panel that allows municipalities and traffic authorities to test Buca's digital twin and see the results in real-time. This panel will be used before decisions such as road closures, traffic changes, or signalization modifications are implemented in the physical environment [10].

## References

1. Braess, D.: Über die Paradoxie des Verkehrs, *Networks*, 1(2), 112–117 (1969).
2. TomTom: Traffic API – Flow Segment Data, TomTom Developer Documentation (2024).
3. Google: Roads API – Speed Limits & Snapped Roads, Google Maps Platform Documentation (2024).
4. Behrisch, M. et al.: SUMO – Simulation of Urban Mobility, in The Third International Conference on Advances in System Simulation (SIMUL 2011), Barcelona, Spain (2011).
5. Karayolları Genel Müdürlüğü (KGM): Trafik İşaretleri El Kitabı, Ankara, KGM Yayınları (2015).
6. "Trafik İşaretleri Hakkında Yönetmelik," T.C. Resmi Gazete, Ankara, Türkiye.

7. Federal Highway Administration (FHWA): Manual on Uniform Traffic Control Devices (MUTCD), Part 4: Highway Traffic Signals, Washington, DC: U.S. Dept. of Transportation (2009).
8. Federal Highway Administration (FHWA): Traffic Signal Timing Manual, 2nd ed., Washington, DC: U.S. Dept. of Transportation (2008).
9. Trafik Kontrol Donanımları - Sinyal Lambaları, TS EN 12368, Türk Standartları Enstitüsü (TSE), Ankara (Türkiye).
10. T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı: İstanbul Adaptif Trafik Yönetim Sistemi (ATAK) (2019).
11. ISSD: Sinyalizasyon Gereklilik Analizi, ISSD Akademi (2024).
12. Burov, M.: Detecting Braess Routes an Algorithm Accounting (2021).

# Computer Simulation of the Elastic Regime of the Development of a Single-Well Reservoir with an Unknown Initial State

Khanlar Gamzaev[0000-0002-1228-7892] and Sevinc Kerimova
*Azerbaijan State Oil and Industry University*
Baku, Azerbaijan
gamzaevkhanlar@gmail.com
kerimovasevinc66@gmail.com

**Abstract.** To describe the elastic regime of the development of a single-well oil reservoir, a model of a turbulent filtration flow of a single-phase liquid is proposed, presented in the form of a nonlinear parabolic equation with respect to the volume flow of the liquid. The laws of change in the time of the well flow rate and the pressure at the bottom of the well are considered to be set. The distribution of fluid flow in the reservoir at the initial moment of time, as well as the presence of a hydrodynamic connection between the reservoir and the external region, are considered unknown. Within the framework of the proposed model, the task is to determine the distribution of fluid flow in the reservoir, including the boundary. This problem belongs to the class of boundary value inverse problems without initial conditions. To numerically solve the inverse problem, a discrete analogue of the problem on a uniform difference grid is first constructed using the method of difference approximation. To solve the resulting system of linear equations, a computational algorithm is proposed that makes it possible to determine the flow rate of a liquid at all nodal points of the calculated part, with the exception of a certain triangular area. The calculated value of the fluid flow rate at the outer boundary of the formation determines the presence or absence of a hydrodynamic connection between the formation and the outer region. Numerical experiments for a model oil reservoir were carried out based on the proposed computational algorithm.

**Keywords:** single-well oil reservoir, elastic development mode, turbulent filtration law, boundary value inverse problem without initial conditions, difference approximation method.

## 1 Introduction

It is known that in the initial period of oil field development, reservoir pressure significantly exceeds the pressure of oil saturation with gas and a natural elastic development regime occurs in the reservoir. In the elastic mode of development, oil inflow to wells occurs due to the use of energy from the elastic expansion of the oil itself and the formation rock. Currently, a mathematical model of the unsteady filtration flow of a single-phase liquid in a reservoir is mainly used to analyze the elastic regime of an oil reservoir and forecast development indicators. This model

includes the differential equation of continuity of a single-phase filtration flow, the equation of fluid flow motion, which uses one or another empirical filtration law, and additional equations of state for a liquid and a porous medium [1-3]. In practice, when modeling a single-phase filtration flow in a reservoir, Darcy's law is mainly used as the equation of motion, expressing the linear dependence of the filtration rate on the pressure gradient. However, Darcy's law applies only to Newtonian fluids in a certain limited range of filtration rates, in which turbulence, inertia, and other high-velocity effects are negligible. Numerous experimental studies show that as the filtration rate increases, first due to inertial effects, and then due to turbulence in the fluid flow, a deviation from Darcy's law occurs. It is believed that this range of filtration rates can be adequately represented by the so-called law of turbulent filtration, which describes both laminar flow and flow with inertial-turbulent effects in the reservoir [2–4].

Obviously, in order to solve practical problems related to the development of reservoirs in an elastic regime, in addition to the mathematical model, it is necessary to know the initial and boundary conditions describing the initial state of the reservoir and the interaction of the reservoir with its surrounding area. However, it should be noted that access to the oil reservoir is limited and is possible only through wells opening the reservoir in small areas, the initial pressure distribution in the reservoir, as well as the pressure and flow of oil at the outer boundary of the reservoir are not available for direct measurements. Therefore, it is practically impossible to accurately represent the initial conditions in the formation and the conditions at the outer boundary of the formation. The main sources of information about the processes occurring in the reservoir during development are production wells, where the pressure and flow rate of the well are available for direct measurements. In this regard, for the practice of developing oil reservoirs in an elastic regime, it is important to determine development indicators based only on information obtained from wells.

In this paper, we study the issue of numerical modeling of the elastic regime of the development of a single-well reservoir using the law of turbulent filtration, in the absence of an initial condition and a condition at the outer boundary of the reservoir.

## 2 Problem statement and solution method

Let us consider a homogeneous cylindrical oil-bearing reservoir with a thickness $h$ and radius $R$, bounded from above and below by impenetrable planes. In the center of the cylindrical reservoir there is a hydro dynamically perfect borehole of radius $r_w$, the axis of which coincides with the axis of the cylinder. The pressure at the bottom of the well exceeds the pressure of oil saturation with gas and the reservoir is developed in an elastic mode. It is assumed that the inflow of oil from the reservoir to the well is a plane-radial flow of a single-phase liquid obeying the law of turbulent filtration. Considering oil to be a weakly compressible viscous liquid, we will present a mathematical model of the elastic regime of the development of this reservoir in the form of a system of equations that includes the differential equation of continuity of a single-phase filtration flow

$$\beta^* \frac{\partial P(r,t)}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(ru(r,t)\right), \tag{1}$$

and the equations of motion of the filtration flow, which uses the law of turbulent filtration

$$\frac{\partial P(r,t)}{\partial r} = \frac{\mu}{k}u(r,t) + \beta\rho u^2(r,t), \tag{2}$$

where $P(r,t)$ is the pressure, $u(r,t)$ is the filtration rate of the liquid (oil), $\rho$ is the density of the liquid, $\mu$ is the dynamic viscosity of the liquid, $k$ is the absolute permeability of the porous medium, $\beta^*$ is the elastic capacity coefficient of the reservoir, $\beta$ is the turbulence coefficient.

Taking into account the relationship between the filtration rate $u(r,t)$ and the volumetric flow rate of liquid $Q(r,t)$, we represent the systems of equations (1) and (2) in the form

$$\frac{\partial P(r,t)}{\partial t} = \frac{1}{2\pi r h\beta^*}\frac{\partial Q(r,t)}{\partial r}, \tag{3}$$

$$\frac{\partial P(r,t)}{\partial r} = \frac{\mu}{2\pi r h k}Q(r,t) + \frac{1}{(2\pi r h)^2}\beta\rho Q^2(r,t). \tag{4}$$

The systems of equations (3) and (4) can be reduced to a single equation with respect to the volume flow of a liquid $Q(r,t)$. To do this, we differentiate both parts of equation (4) by the variable $t$

$$\frac{\partial^2 P(r,t)}{\partial r\partial t} = \left(\frac{\mu}{2\pi r h k} + \frac{\beta\rho}{2(\pi r h)^2}Q(r,t)\right)\frac{\partial Q(r,t)}{\partial t}.$$

Substituting the expression $\frac{\partial P(r,t)}{\partial t}$ from equation (3) here, we obtain a nonlinear parabolic equation with respect to the function $Q(r,t)$

$$\left(\frac{\mu\beta^*}{k} + \frac{1}{\pi r h}\beta\beta^*\rho Q(r,t)\right)\frac{\partial Q(r,t)}{\partial t} = \frac{\partial^2 Q(r,t)}{\partial r^2} - \frac{1}{r}\frac{\partial Q(r,t)}{\partial r}. \tag{5}$$

Obviously, in order to determine the distribution of the volume flow of a liquid in the elastic mode of development based on model (3), it is necessary to have initial and boundary conditions describing the initial state of the reservoir and its interaction with the surrounding area.

Since the flow rate of the well and the pressure at the bottom of the well are available for direct measurements, the boundary conditions at the well can be represented as

$$Q(r_w,t) = Q_w(t), \tag{6}$$

$$P(r_w,t) = f(t), \tag{7}$$

where $Q_w(t)$ and $f(t)$ are the specified functions.

However, due to the fact that the initial distribution of fluid flow in the formation and the fluid flow at the outer boundary of the formation $r = R$ are not directly measurable, it is not possible to formulate the initial condition and boundary condition corresponding to the interaction of the formation with its surrounding area. Using the equation of continuity of the filtration flow (3) and the boundary condition with respect to pressure (7), we obtain an additional condition with respect to fluid flow in the well

$$\frac{\partial Q(r_w,t)}{\partial r} = 2\pi r_w h\beta^* \frac{df(t)}{dt}. \tag{8}$$

Thus, the task of determining the dynamics of the distribution of the volume flow of liquid in the reservoir under consideration is reduced to solving equation (5) under conditions (6), (8). The assigned task (5), (6), (8) it belongs to the class of boundary-value inverse problems without initial conditions [5-8]. Numerous works have been devoted to theoretical research and the development of numerical methods for solving inverse problems related to the restoration of boundary conditions [5, 6, 9-18].

For the numerical solution of the inverse problem (5), (6), (8), first, we discretize a given rectangular area $\{r_w \le r \le R,\ \ 0 < t \le T\}$ in space and time, with steps $\Delta r = \frac{R - r_w}{n}$ in the variable $r$ and $\Delta t = \frac{T}{m}$ in the variable $t$

$$\overline{\omega} = \left\{(r_i, t_j) : r_i = r_w + i\Delta r,\ t_j = j\Delta t,\ i = 0,1,2,...n,\ j = 0,1,2,...m\right\}.$$

Let's build a discrete model of the problem (5), (6), (8) on a grid $\overline{\omega}$ using the difference approximation method. Using an explicit time approximation for the nonlinear coefficient in equation (5), we present a discrete model of the problem as

$$\left(\frac{\mu\beta^*}{k} + \frac{\rho\beta\beta^*}{\pi r_i h} Q_i^{j-1}\right)\frac{Q_i^j - Q_i^{j-1}}{\Delta t} = \frac{Q_{i+1}^j - 2Q_i^j + Q_{i-1}^j}{\Delta r^2} - \frac{1}{r_i}\frac{Q_i^j - Q_{i-1}^j}{\Delta r}, \tag{9}$$

$$i = \overline{1, n-1}, \qquad j = \overline{1, m},$$

$$Q_0^j = Q_w{}^j, \qquad j = \overline{0, m}, \tag{10}$$

$$\frac{Q_1^j - Q_0^j}{\Delta r} = 2\pi r_w h\beta^* \frac{f^j - f^{j-1}}{\Delta t},\ j = \overline{0, m}, \tag{11}$$

where $Q_i^j \approx Q(x_i, t_j)$, $f^j = f(t_j)$, $Q_w{}^j = Q_w(t_j)$.

As you can see, the discrete model of the problem (5), (6), (8) for each fixed value $j$, $j = 1, .2 \ldots, m$, it is a system of linear difference equations in which the unknowns $Q_i^j$, $i = 1, 2, ...n,$ are the approximate values of the desired function $Q(r,t)$ at the nodal points of the difference grid $\overline{\omega}$.

According to conditions (10), (11) on two spatial layers $r_0 = r_w$, $r_1 = r_0 + \Delta r$ the values of the variables $Q_0^j$ and $Q_1^j$ are considered to be set (in the figure, the corresponding nodes are indicated by circles), i.e.

$$Q_0^j = Q_w^j, \qquad Q_1^j = Q_0^j + 2\pi r_w h\beta^* \Delta r \frac{f^j - f^{j-1}}{\Delta t}.$$

Now the task is to find from the system of equations (9) the approximate values of the desired function $Q(r,t)$ at the remaining nodal points, i.e. $Q_i^j$, $i = \overline{2,n}$, $j = \overline{0,m}$.

A simple analysis of the system of equations (9) shows that from this system it is consistently possible to determine:

$$Q_2^j, \text{ for } i=1, \ j = \overline{1,m};$$
$$Q_3^j, \text{ for } i=2, \ j = \overline{2,m};$$
$$Q_4^j, \text{ for } i=3, \ j = \overline{3,m};$$
$$Q_5^j, \text{ for } i=4, \ j = \overline{4,m};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Q_n^j, \text{ for } i=n-1, \ j = \overline{n-1,m}.$$

In the figure, the grid nodes corresponding to the calculated variables are indicated by asterisks. However, as shown in the figure, in some nodes located in the triangular region D, the values of the desired function are not determined.

The numerical implementation of the above calculation procedure when setting $Q_0^j$ and $Q_1^j$ can be carried out using the recurrent formula

$$Q_{i+1}^j = \Delta r^2 \left(\frac{\mu\beta^*}{k} + \frac{1}{\pi r_i h}\beta\beta^*\rho Q_i^{j-1}\right)\frac{Q_i^j - Q_i^{j-1}}{\Delta t} + \frac{\Delta r\left(Q_i^j - Q_{i-1}^j\right)}{r_i} + 2Q_i^j - Q_{i-1}^j, \qquad (12)$$
$$i = \overline{1,n-1}, \quad j = \overline{i,m}.$$

Thus, formula (12) allows us to calculate approximate values of the desired function $Q(r,t)$ at the nodal points of the difference grid, including nodes on the outer boundary of the reservoir, with the exception of unmarked nodes in the D region.

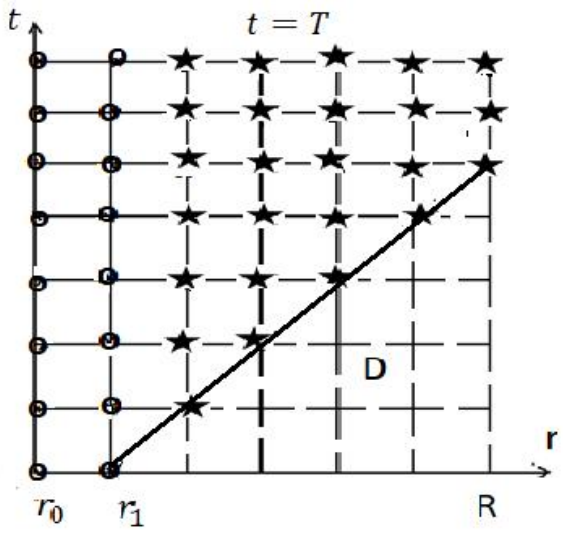



Fig. Discrete domain of numerical solution of the inverse problem

According to the value of the fluid flow rate $Q_n^j$, $j = \overline{n-1,m}$, at the outer boundary, the presence or absence of a hydrodynamic connection between the reservoir and the outer region is determined.

## 3 Results of numerical experiments

Based on the proposed computational algorithm, numerical experiments were conducted for a model single-well oil reservoir with the following characteristics:
reservoir radius $R$ =100 m;
layer thickness $h$ =10 m;
the coefficient of absolute permeability of the formation $k = 5 \cdot 10^{-12}$ m$^2$;
well flow rate $Q_w(t)$ = 120 m$^3$/day;
dynamic viscosity of oil $\mu = 3 \cdot 10^{-3}$ Pa · s;
turbulence coefficient $\beta$ =10$^6$ m$^{-1}$;
oil density $\rho$ = 850 kg/m$^3$;
the coefficient of elastic capacity of the reservoir $\beta^* = 2 \cdot 10^{-9}$ Pa$^{-1}$.

Numerical experiments were carried out according to the following scheme:
I. The distribution of the volume flow of liquid in the reservoir at the initial moment of time, the flow rate of the well $Q_w(t)$ and the volume flow of liquid at the outer boundary of the reservoir $Q_R(t)$ are set. A numerical solution of equation (5) satisfying a given initial condition and boundary conditions

$$Q(r_w,t) = Q_w(t), \qquad Q(R,t) = Q_R(t)$$

is determined. In other words, a direct problem is solved numerically to determine the values of the function $Q(r,t)$ at the nodal points of the difference grid $\overline{\omega}$.

II. In the course of solving the direct problem, a variable $\frac{df(t)}{dt}$ is determined from relation (8) and the resulting dependence $\frac{\partial Q(r_w,t)}{\partial r} = 2\pi r_w h\beta^* \frac{df(t)}{dt}$ is taken as an additional condition for solving the inverse problem. (5), (6), (8) to restore the volume flow of liquid at the outer boundary of the reservoir.

According to the proposed scheme, numerical experiments were carried out to restore two functions $Q_R(t) = \frac{0.02}{\sqrt{t}}$; $Q_R(t) = 0.0$. At the same time, the self-regularization method (natural regularization method) is used to find a stable solution to the inverse problem [5, 6]. The time discretization step $\Delta t$ is chosen as the regularization parameter. The calculations were performed on a uniform space-time difference grid with steps of $\Delta r = 3.996$ m, $\Delta t = 30$; 60; 120 s. Numerical experiments revealed that with small sampling steps in time ($\Delta t < 30$ s), an unstable solution appears and the volumetric flow rate of the liquid at the outer boundary of the reservoir is determined by a very large error. However, as the sampling step ($\Delta t \geq 30$ s) increases, the accuracy of reconstructing the volumetric flow rate of the liquid at the boundary increases. The results of numerical experiments conducted at $\Delta t = 60$ s are presented in the table. It contains $t$ –the time, $r$ – the distance from the center of the well, the exact $\overline{Q}$ and calculated values $\widetilde{Q}$ of the function $Q(r,t)$. The dynamics of the distribution of the volume flow of liquid in the reservoir, when the above functions are restored, are presented in columns 2-5 and 6-9, respectively.

Table. Calculated distributions of fluid flow in the reservoir

| $r$, m | $\overline{Q}\times10^3$, $\widetilde{Q}\times10^3$ m³/s | | | | $\overline{Q}\times10^3$, $\widetilde{Q}\times10^3$ m³/s | | | |
|---|---|---|---|---|---|---|---|---|
| | $t$=10 | $t$=30 | $t$=50 | $t$=60 | $t$=10 | $t$=30 | $t$=50 | $t$=60 |
| 0.100 | 1.39 | 1.39 | 1.39 | 1.39 | 1.39 | 1.39 | 1.39 | 1.39 |
| 4.096 | 1.36 | 1.38 | 1.39 | 1.39 | 1.36 | 1.38 | 1.38 | 1.38 |
| 8.092 | 1.32 | 1.37 | 1.38 | 1.38 | 1.32 | 1.37 | 1.37 | 1.37 |
| 12.088 | 1.25 | 1.35 | 1.37 | 1.37 | 1.25 | 1.34 | 1.36 | 1.36 |
| 16.084 | 1.17 | 1.33 | 1.35 | 1.36 | 1.17 | 1.31 | 1.34 | 1.34 |
| 20.080 | 1.08 | 1.31 | 1.34 | 1.34 | 1.07 | 1.28 | 1.31 | 1.32 |
| 24.076 | | 1.27 | 1.32 | 1.32 | | 1.24 | 1.28 | 1.29 |
| 28.072 | | 1.24 | 1.29 | 1.30 | | 1.19 | 1.25 | 1.26 |
| 32.068 | | 1.20 | 1.26 | 1.27 | | 1.14 | 1.21 | 1.22 |
| 36.064 | | 1.16 | 1.23 | 1.24 | | 1.08 | 1.17 | 1.18 |
| 40.060 | | 1.12 | 1.20 | 1.21 | | 1.02 | 1.12 | 1.14 |
| 44.056 | | 1.08 | 1.17 | 1.17 | | 0.96 | 1.07 | 1.09 |
| 48.052 | | 1.03 | 1.13 | 1.13 | | 0.90 | 1.01 | 1.03 |
| 52.048 | | 0.99 | 1.09 | 1.09 | | 0.83 | 0.96 | 0.98 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 56.044 | | 0.95 | 1.04 | 1.05 | | 0.76 | 0.89 | 0.92 |
| 60.040 | | 0.91 | 0.99 | 1.00 | | 0.70 | 0.83 | 0.85 |
| 64.036 | | | 0.95 | 0.95 | | | 0.76 | 0.78 |
| 68.032 | | | 0.89 | 0.89 | | | 0.69 | 0.71 |
| 72.028 | | | 0.84 | 0.84 | | | 0.61 | 0.64 |
| 76.024 | | | 0.78 | 0.78 | | | 0.54 | 0.56 |
| 80.020 | | | 0.72 | 0.71 | | | 0.46 | 0.47 |
| 84.016 | | | 0.66 | 0.64 | | | 0.37 | 0.39 |
| 88.012 | | | 0.59 | 0.57 | | | 0.24 | 0.30 |
| 92.008 | | | 0.52 | 0.50 | | | 0.19 | 0.20 |
| 96.004 | | | 0.44 | 0.42 | | | 0.10 | 0.10 |
| 100.00 | | | 0.37 | 0.33 | | | 0.00 | 0.00 |

The results of numerical experiments show that with the beginning of reservoir development, the distribution of the volume flow rate of the liquid is restored in a time-varying region. For example, 30 minutes after the start of reservoir development, the distribution of fluid flow is restored within a radius of 60.04 meters around the well. And 50 minutes after the start of development, the fluid flow distribution is fully restored in the reservoir.

Analysis of the results obtained shows that in both variants, the volumetric flow rate of the liquid at the outer boundary of the formation, as well as the dynamics of the distribution of the volumetric flow rate of the liquid in the formation, with the exception of the specified area D, is restored absolutely accurately. Therefore, the exact and calculated distribution of fluid flow in the reservoir are presented in the same column of the table.

## 4 Conclusion

The problem of determining the conditions at the outer boundary of a single-well oil reservoir developed in an elastic regime is considered. The mathematical model of this problem, based on the law of turbulent filtration, is presented as a boundary-value inverse problem without initial conditions for a nonlinear parabolic equation with respect to the volume flow of a liquid. The proposed computational algorithm allows, based only on the information received from the well, to determine the presence or absence of a hydrodynamic connection between the reservoir and the surrounding area, as well as the dynamics of the distribution of fluid flow during the development of the reservoir in an elastic mode.

## References

1. Shchelkachev, V.N., Lapuk, B.B.: Underground Hydraulics. Izhevsk, Regular and chaotic dynamics (2001).
2. Basniev, K. S., Dmitriev, N. M. Rozenberg, G.D.: Oil and Gas Hydromechanics. Izhevsk, Institute of Computer Research (2005).
3. Vishnyakov, V.V., Suleimanov, B.A., Salmanov, A.V., Zeynalov, E.B.: Primer on enhanced oil recovery. Gulf Professional Publishing (2019).
4. Aziz, K., Settari, A.: Petroleum reservoir simulation. New York, Applied Science Publishers ( 1979).
5. Alifanov, O. M.: Inverse heat transfer problems. Berlin, Springer ( 2011).

6. Samarskii, A.A., Vabishchevich, P.N.: Numerical methods for solving inverse problems of mathematical physics. Berlin, De Gruyter (2008).
7. Yaparova, N.M.: Method for solving some multidimensional inverse boundary value problems for parabolic PDEs without initial conditions. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics 15( 2), 97–108 (2015).
8. Gamzaev, Kh. M.: Difference Method of Solving an Inverse Problem for the convective diffusion equation. Journal of Engineering Physics and Thermo-physics 84 (3), 526–532 (2011).
9. Prilepko, A. I., Orlovsky, D. G., Vasin, I. A.: Methods for solving inverse problems in mathematical physics. New York, Marcel Dekker (2000).
10. Kozhanov, A.I.: Inverse Problems for Determining Boundary Regimes for Some Equations of Sobolev Type. Bulletin of the South Ural State University, Ser. Mathematical Modelling, Programming & Computer Software 9(2), 37–45 ( 2016).
11. Tabarintseva, E. V.: On the estimate of accuracy of the auxiliary boundary conditions method for solving a boundary value inverse problem for a nonlinear equation. Numerical Analysis and Applications 11(3), 236–255 (2018).
12. Dmitriev, V.I., Stolyarov, L.V.: Numerical method for the inverse boundary-value problem of the heat equation. Computational Mathematics and Modeling 28(2),141–147 (2017).
13. Sidikova, A.I.: A Study of an Inverse Boundary Value Problem for the Heat Conduction Equation. Numerical Analysis and Applications 12, 70–86 ( 2019).
14. Gamzaev, Kh. M.: Numerical identification of hydrodynamic parameters of a reservoir under elastic-water-drive development mode. Bulletin of the South Ural StateUniversity. Ser. Mathematical Modelling, Programming & Computer Software 18(4), 56–65 (2025).
15. Denisov, A. M.: Approximate solution of inverse problems for the heat equation with a singular perturbation. Computational Mathematics and Mathematical Physics 61(12), 2004–2014( 2021).
16. Yaparova, N.M.: Numerical Methods for Solving a Boundary Value Inverse Heat Conduction Problem. Inverse Problems in Science and Engineering 22(5), 832-847 (2014).
17. Gamzaev, Kh. M.: Identification of the boundary regime in the process of water- oil displacement from the reservoir. St. Petersburg Polytechnic University. Journal: Physics and Mathematics 17(4), 57–67 (2024).
18. Koleva, Miglena N., Vulkov, Lubin. G.: Computational Solution of an Inverse Boundary-Value Problem for Heat Transfer in a Composite Material. Applied Sciences 15(18), 10230 (2025).

# Impact of Optimizer: Comparative Analysis of Loss Functions in Different Models

Abdulvahit Karail[1][0009000387188906] and Yasin Ortakcı[2][0000-0002-0683-2049]

[1] Karabuk University, Vocational School of Information Technologies, Karabük, Türkiye
[2] Karabuk University, Dept. of Computer Engineering, Karabük, Türkiye
abdulvahitkarail@karabuk.edu.tr

**Abstract.** This study compares the performance of different artificial neural network models for data classification problems. The models used include Multilayer Perceptron (MLP), Kolmogorov–Arnold Network (KAN), and Liquid Time-Constant Networks (LTC). Model training was performed on the MITBIH ECG dataset, and modern optimization strategies with different loss functions were systematically compared. Evaluation metrics such as accuracy, macro-F1, precision, and recall were assessed. The findings show that KAN models provide high accuracy, while LTC models operate more slowly due to their time-step based calculations.

**Keywords:** Optimizers, loss functions, Kolmogorov–Arnold Network, Liquid Time-Constant Networks, Multilayer Perceptron.

## 1 Introduction

A significant gap in the current literature is the absence of an evidence-based implementation guide that evaluates the interactions between loss functions and optimizer for artificial neural network (ANN), particularly when processing unstable health data such as ECGs. This study was conducted because determining the compatibility of a particular loss function and optimizer with the applied model is a challenging task.

Determining the optimal fit between a given loss function, an optimizer, and the implemented model remains a challenging task. To overcome this challenge, we conducted a systematic and reproducible experimental study on the MIT-BIH Arrhythmia dataset. This study comparatively examines modern artificial neural network models for solving the data classification problem. Three different model families were selected for the experimental analysis. These models are Multilayer Perceptron (MLP), Kolmogorov Arnold Network (KAN), and Liquid Time Constant Networks (LTC). A comprehensive experimental design was implemented using various loss functions such as Cross Entropy (CE), Focal Loss (FL) and Label Smoothing Cross Entropy (LSCE), optimizer such as Adam, AdamW, Stochastic

Gradient Descent (SGD), Adagrad and RMSProp. Each model was tested with 3 different loss functions and 5 different optimizers. Thus, the study revealed which architecture was more suitable in terms of both performance and computational cost. The results we have obtained are practice-oriented guidance.

In comparative experiments, CE/LSCE consistently provided high accuracy in imbalanced ECG data when used in conjunction with Adam/AdamW. KAN showed competitive performance with MLP and under the evaluated representation and time step settings, LTC was observed to train more slowly. These observations shed light on the suitability of the model training strategy under data unstable and computational constraints.

The geometry of the loss surface, local minima, saddles, straight directions, multiple fractal structure, and large-scale connectivity, fundamentally shapes the optimization trajectories, convergence speed, and generalization performance across model families. Therefore, a comparative study aiming to reveal how different loss functions interact with the optimization landscapes of KAN, MLP, and LTC should utilize diagnostics that capture the curvature, modality, and dynamic signatures of the optimization, and select loss targets and optimization algorithms whose interactions with the landscape structure are well-defined in literature [1-4].

Understanding the interaction between the choice of loss function and the resulting optimizers is crucial for predicting training dynamics and the ability to generalize across model families [3], [4], [5].

This study fills a gap in the literature by systematically comparing KAN, MLP, and LTC architectures on the MIT-BIH Arrhythmia dataset under a single documented and reproducible protocol using three loss-based and five optimizers. The evaluation focuses not only on accuracy but also on practical metrics such as macro-F1.

The rest of the article is structured as follows. Related work in section 2 summarizes previous work and identifies research gaps. Section 3 examines preprocessing and details the architecture of ANN, loss functions and optimizers. Section 4 examines the methodology. Section 5 describes experimental studies and analyzes the results. Section 6 summarizes the conclusions.

## 2 Related Work

While prior work covers architecture, loss functions, and optimizers, joint comparisons of MLP/KAN/LTC, CE/Focal Loss/LSCE, SGD/Adam/AdamW/RMSProp/Adagrad under a single, documented protocol on MIT-BIH, with accuracy, macro-F1 reported together, remain limited. Our study addresses this gap and offers practice oriented guidance for ECG classification under class imbalance and compute constraints.

Prior work provides comprehensive surveys of MLP architecture, training, regularization/generalization, design choices, variants and hybrids, and application domains. The synthesis integrates empirical findings and methodological recommendations reported in applied and methodological studies, creating a

consistent guide for researchers and practitioners who design, train, evaluate or compare MLPs in applied settings [6-13].

KAN variants have been used for rolling bearing fault diagnosis, EEG seizure detection, and ECG arrhythmia classification, reporting improved robustness and interpretability in those settings [14], [15].

Training LTCs benefits from principled loss selection and from using adaptive gradient optimizers together with solver–optimizer co design to stabilize gradients and improve convergence [16-22].

Cross entropy is a canonical supervised classification loss that compares a predicted probability distribution to a target distribution and yields gradients suitable for gradient based training. Once a model produces a differentiable output, a differentiable loss such as cross entropy may be used to train the network by gradient based procedures. Wunderlich & Pehle show that exact gradients can be computed for a general loss function, which includes cross entropy as a standard differentiable classification loss. Wunderlich and Pehle show that exact gradients can be obtained for a general differentiable loss and because Yin et al. demonstrate training schemes for liquid neurons compatible with differentiable losses, any differentiable variant of cross entropy is theoretically trainable in LTC/spiking frameworks using the methods described in [23], [24].

Multiple empirical studies document that adaptive optimizers such as Adam, RMSProp, Adagrad typically enable faster decrease of training loss and faster wall clock progression during early epochs, whereas SGD with momentum tends to produce solutions with better generalization when carefully tuned, this tradeoff is widely reported in both survey and empirical benchmark studies [25], [26], [28], [29], [30].

Modern optimization algorithm selection requires striking a pragmatic balance between speed, stability, and generalization. Adaptive optimizer such as Adam and RMSProp should be used when rapid development and ease of adjustment are important, AdamW should be used when weight reduction adjustment is required, Adagrad should be preferred for sparse problems, and SGD or a hybrid adaptive SGD strategy should be used when ultimate generalization performance is prioritized. The aforementioned survey, robustness, and comparative studies provide both the theoretical framework and empirical evidence that motivate these recommendations [25], [26], [27], [28], [30].

Precision penalizes false positives while Recall penalizes false negatives. The operating decision threshold on continuous prediction scores controls this trade-off, and practitioners inspect Precision–Recall pairs and the confusion matrix to select thresholds appropriate to their risk profile, many medical and detection studies report sensitivity and specificity alongside precision and accuracy to characterize clinically relevant trade-offs [32], [33], [34].

F1-Score balances precision and recall and is particularly useful when a single composite metric is required and the costs of false positives and false negatives are comparable. Applied classification studies often report both per-class F1 and per-class precision/recall so that readers can examine asymmetric error patterns [31], [32].

# 3 Background

Loss functions form the training dynamics of neural networks. A loss function maps the predicted outputs and the actual value to a scalar that quantifies the model error and minimizing the expected loss on the data is the fundamental optimization goal of learning.
Optimizer uses the gradients of this objective to update parameters to reduce loss on the training data. Optimizer are methods for finding the minimum or maximum values of objective functions. The selection and configuration of an optimization algorithm greatly influence the quality of the solution, the speed of convergence, the ability to generalize, and the computational cost.

## 3.1 Loss Functions

To analyze the effects of the KAN, LTC and MLP models on optimization, three different loss functions were compared. Cross-Entropy Loss, Weighted Focal Loss and Label Smoothing Cross-Entropy were used as loss functions.
In classification problems, the CE standard approach minimizes the difference between the predicted probability distribution and the actual labels. The multi-class CE loss used in this study is defined in Eq. (1):

$$L_{CE} = -\sum_{c=1}^{C} y_c \log(p_c) \tag{1}$$

To reduce class imbalance and the dominance of easy examples, Focal loss is designed to down weight examples and focus learning on hard examples. It is formulated by adding a modulation factor. The formulation of Focal Loss is given in Eq. (2):

$$L_{FL} = -(1 - p_t)^{\gamma} log(p_t) \tag{2}$$

Here, $p_t$ is the probability that the model predicts for the correct class. It is commonly expressed as a modulated version of cross entropy with a tunable focusing parameter γ and an optional class weighting factor α to address class imbalance.

Label Smoothing is a function used in ANN models to prevent overconfidence and improve the model's generalization ability. To break the model's overconfidence and increase its generalization ability, softened labels are used instead of hard labels. The smoothed target obtained by label smoothing is defined in Eq. (3):

$$y_c^{LS} = (1 - \epsilon) y_c + \frac{\epsilon}{C} \tag{3}$$

Here, ϵ is smoothing factor.

## 3.2 Optimizer

The general structure of an optimizer consists of the basic steps of gradient

calculation, internal state update, direction and magnitude determination, and parameter update. It is expressed as a mathematical formula as follow Eq. (4):

$$w_{t+1} = w_t - \eta \cdot \phi(g_t, M_t) \tag{4}$$

Here, w is model parameters, weights, $\eta$ is learning rate, $g_t$ is current gradient, $M_t$ is optimizer's memory and $\phi$ is optimizer's decision function.

Adam is a popular optimizer that combines both momentum and adaptive learning rate methods. For each parameter, the average and the average of the squares of past gradients are maintained. Parameter update for Adam is given as Eq. (5):

$$\theta_{t+1} = \theta_t - \alpha \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}}+\varepsilon} \tag{5}$$

Here, $\alpha$ is learning rate, $\widehat{m_t}$ is first moment of gradients, $v_t$ is second moment of gradient squares, ε is small constant.

AdamW is a variant of the Adam algorithm that implements weight decay more accurately. While weight decay in classical Adam is implemented by adding it to gradients, AdamW performs it separately from parameter updates. Parameter update for AdamW is given as Eq. (6):

$$\theta_{t+1} = \theta_t - \eta \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}}+\varepsilon} - \eta \lambda \theta_t \tag{6}$$

Here, η is learning rate, λ is weight decay constant.

SGD is one of the most basic optimizer. It calculates the gradient and updates parameters using a small subset of data in each iteration. Parameter update for SGD is given as Eq. (7):

$$\theta_{t+1,i} = \theta_{t,i} - \eta\, g_{t,i} \tag{7}$$

Here, $\theta_t$ is the vector of all parameters in t-th iteration, $\theta_{t,i}$ is the scalar value of the i-th parameter at the t-th iteration, $g_{t,i}$ is the i-th component of the gradient.

RMSProp is an algorithm that provides an adaptive learning rate. By keeping the exponential mean of the squares of the gradients, it determines a different learning rate for each parameter. Parameter update for RMSProp is given as Eq. (8):

$$\theta_{t+1} = \theta_t - \eta \frac{g_t}{\sqrt{E[g_t^2]}+\varepsilon} \tag{8}$$

Here, ρ is decay rate, $\theta_t$ is parameters in step t, $g_t = \nabla_{\theta_t}\mathcal{L}_t$ is the gradient of that step.

Adagrad is an algorithm that adapts the learning rate based on the sum of the squares of the past gradients for each parameter. It reduces the learning rate for frequently updated parameters while keeping the learning rate high for infrequently updated parameters. Parameter update each component for Adagrad is given as Eq.

(9):

$$\theta_{t+1,i} = \theta_{t,i} - \eta \frac{g_{t,i}}{\sqrt{G_{t,i}+\varepsilon}} \quad (9)$$

Here, i is parameter index and $G_{(t,i)}$ is sum of the squared gradients of the parameter.

### 3.3 Architecture of ANN Models

#### 3.3.1 Multilayer Perceptron

MLP is a classic artificial neural network consisting of fully connected layers. It is one of the most fundamental architectures of artificial neural networks and is composed of fully connected layers. In each layer, the input data undergoes linear transformation, followed by the application of nonlinear activation functions. This structure is suitable for learning relationships between features and is widely used, particularly in data classification. Its simplicity and rapid training make it a preferred choice in many applications. Figure 1 shows the structure of MLP.

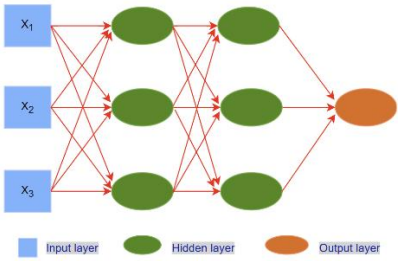



**Fig. 1.** Structure of MLP

#### 3.3.2 Kolmogorov Arnold Network

The Kolmogorov Arnold Network is a new artificial neural network architecture introduced in 2024 by researchers at MIT, Caltech, and others, offering a radical alternative to MLPs, which have been the standard in deep learning for the past 10-15 years. KAN is an architecture that provides functional flexibility through a kernel-based spline approach. KAN is an architecture developed based on the Kolmogorov–Arnold representation theorem. The overall architecture of the Kolmogorov Arnold Networks is shown in Figure 2.

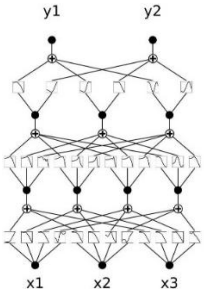



**Fig. 2.** The overall architecture of the Kolmogorov-Arnold Networks [35]

### 3.3.3 Liquid Time-Constant Networks

LTC is a continuous-time dynamic Recurrent Neural Network (RNN)-based model designed for time series modeling. Unlike traditional RNNs, LTC cells are defined by differential equations, and the state of each neuron is continuously updated over time. This feature makes LTC a more flexible and biologically inspired model for time series data. Figure 3 shows the general architecture of Liquid Time Constant Networks.

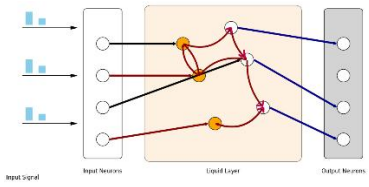


**Fig. 3.** The overall architecture of the Liquid Time-Constant Networks

## 4 Methodology

Our study involved testing three different ANN models using three different loss functions and five different optimizers, resulting in a detailed comparison. Figure 4 shows an overview of the methodology.

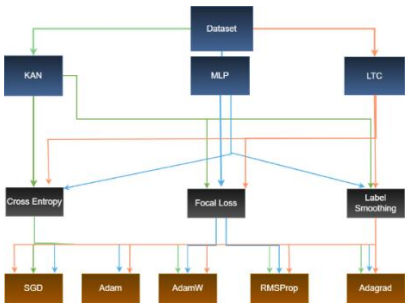



**Fig. 4.** The overview of the methodology

Figure 4 illustrates the end to end experimental stages of the study. Samples from the MIT-BIH Arrhythmia dataset are fed into three different architectures via separate streams after a common preprocessing step. For each architecture, training was conducted using three different loss functions such as Cross-Entropy, Focal Loss, and Label-Smoothing CE and each configuration was run under the same training protocol with five optimizer. This resulted in 45 different outcomes, all combinations of which were evaluated in a comparable and reproducible manner.

In all experiments, data is processed using the same train/validation/test partitioning and Z-score normalization. Hyperparameters such as batch size, epoch, and initial learning rate are kept constant at the baseline level and gradient clipping is applied using the same rules. Class weights for Weighted CE were derived from the training distribution, and alpha and gamma values were reported for Focal Loss. Each architecture-loss-optimizer combination was run with at least one fixed seed, and the results were reported as averages. Thus, convergence speed, stability, and generalization behavior can be isolated from the effects of optimizer and loss selection.

The output of each experiment, accuracy, macro-F1, and precision/recall was aggregated. Colored arrows visually distinguish which architecture was used for training, along with which loss and optimizer. Finally, tables and training validation curves summarize the performance-cost profile of each combination.

Features were scaled using Z-score normalization. For architectures trained with PyTorch, normalization was applied on a sample by sample basis within the dataset. The aim is to prevent data leakage and ensure scale consistency.

### 4.1 Structure of Models

The study compares three main architectures. These are KAN, MLP, and LTC. The MLP architecture used in this study consists of 3 hidden layers by default. Each hidden layer comprises Linear, BatchNorm1d, ReLU, and Dropout blocks, respectively. The network terminates with a linear output layer for classification. The default hidden dimension is 128, and the depth is set to 3, these hyperparameters can be adjusted according to the experimental scenario. The KAN architecture features two KAN layers between the input and output. Each KAN layer applies learnable piecewise linear transformations on a fixed grid for each input feature, and the resulting transformations are summed up along with the bias to produce the output. GELU activation and dropout are used in the intermediate layer. The default settings are grid size 16 and hidden size 64, which can be changed according to requirements. The LTC architecture is built on an LTC cell created using AutoNCP wiring from the NCPs library. The default number of NCP neurons is 12, and hyperparameters such as time step, integrator, and number of neurons can be reconfigured for different setups.

# 5 Experimental studies

In this study, hyperparameters were kept constant to ensure fair and reproducible comparisons. The batch size was set at 128, the initial learning rate at $1\times10^{-3}$, and the training duration at 15 epochs. Additionally, the number of NCP neurons for the LTC architecture was 12. To prevent overlearning during the training process, gradient clipping was applied, and the learning rate scheduler was optionally tested. All experiments were performed in Visual Studio Code using Python.

## 5.1 Data Set

This study utilized the MITBIH Arrhythmia ECG dataset. The dataset consists of time-series data representing different arrhythmia types, commonly used for heart rhythm classification. Each sample represents a heartbeat signal of a specific length and is presented with class labels. With Z-score normalization, the mean and standard deviation were calculated for each feature, thus normalizing the data.

In this study, MIT-BIH Arrhythmia data was loaded from two separate files. These are mitbih train file and mitbih test file. With data splitting, the dataset was divided into 68% training, 12% validation and 20% testing ratios. All samples in the test file were used as the test set. Thus, the data flow was standardized in the form of training, then validation, and finally testing.

## 5.2 Evaluation metrics

In addition to accuracy, macro-F1 was reported as the primary metric to accurately reflect the effect of class imbalance, precision and recall were also measured. Macro F1 was used because it calculates the F1 Score separately for each class in the dataset and then takes the arithmetic mean of these scores, and because the dataset is unbalanced. The primary reason for using these metrics is to evaluate model performance from different perspectives and to avoid the misleading nature of relying on a single measure, such as accuracy, especially in datasets with unbalanced classes.

Accuracy is the ratio of correctly predicted samples to the total sample size. Eq. (10) defines accuracy.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total number of samples}} \tag{10}$$

Precision focuses on reducing false positives. It is critical for lowering the number of false alarms in healthcare applications. Eq. (11) defines precision.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{11}$$

Recall indicates the rate at which true positives are detected. Since missed cases in disease detection can have serious consequences, recall must be high. It shows how

many true positive samples are correctly predicted. Eq. (12) defines Recall.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{12}$$

In unbalanced datasets, F1 is more meaningful than accuracy. It is the harmonic mean of precision and recall. Eq. (13) defines F1-Score.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

### 5.3 Results

Comparisons of training duration and convergence curves with uniform axes/styles Figures 5. Figures 5 show the validation accuracy graphs for the KAN, LTC, MLP.

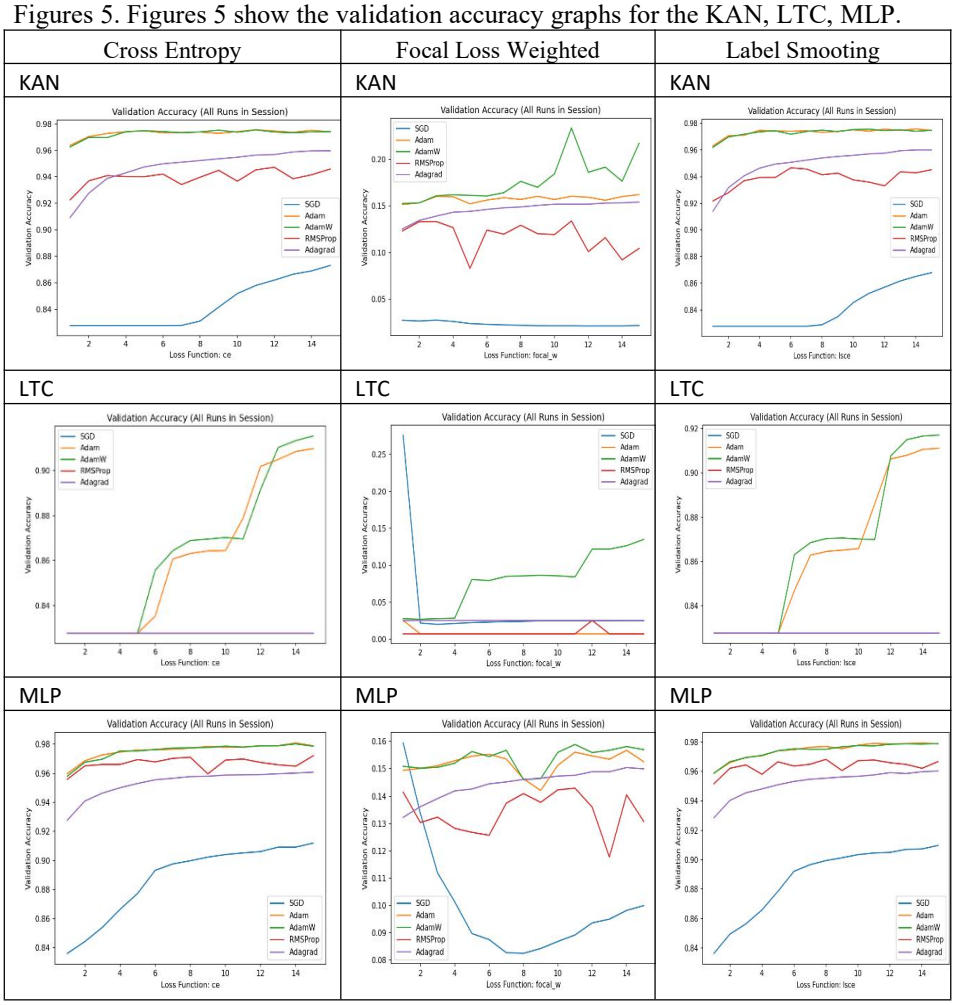



**Fig. 5.** Validation Accuracy Results

When using the KAN model with CE or LSCE and the Adam/AdamW optimizer, fast convergence, high test accuracy, and high macro F1 were achieved. Adagrad, used with the KAN model, was stable but had low minority class performance. RMSProp was choppy and generally lower. SGD was slow and low.

Experiments conducted for MLP show that the choice of optimization is crucial. With Adam/AdamW, both Cross Entropy and LSCE losses converged rapidly within the first few epochs, and validation accuracy reached the 97.5–98% range, this was confirmed in testing with 97.7–97.8% accuracy and approximately 0.87–0.88 macro-F1. RMSProp/Adagrad reached lower plateau values, while SGD showed slow and low convergence without a scheduler/momentum. In Focal Loss configurations, learning did not practically begin due to insufficient α–γ settings, and validation accuracy remained in the 10–16% range. Therefore, the most reliable choice in this dataset is the combination of CE or LSCE and Adam/AdamW.

LSCE and Adam/AdamW yielded stable and high validation accuracy for LTC, label smoothing reduced overconfidence and improved calibration. Learning did not begin with SGD/RMSProp/Adagrad. Results with CE and Adam/AdamW were very close to LSCE, the optimizer was the primary determinant in LTC.

Adam and AdamW have demonstrated the highest accuracy, fastest convergence, and highest stability in all successful scenarios. They are the only working solutions, especially in architectures with complex dynamics like LTC. Adagrad has been a safe but slow alternative to MLP and KAN, but has failed in LTC. RMSProp, on the other hand, has exhibited high volatility and has generally lagged behind the Adam family. SGD proved insufficient for this dataset and architectures within the examined hyperparameter space. The convergence speed is very low, and it failed to overcome local minima.

The results show that F1 values are critical, especially due to class imbalance. Table 1 shows the best configuration such as accuracy, macro-F1 for each model. The best combination results of the models are shown in the appendix Table 1.

**Table 1.** Summary table for the best combinations of models.

| Model | Optimizer | Loss | Test Accuracy | Macro F1 |
|---|---|---|---|---|
| MLP | Adam | CE | 0.978 | 0.879 |
| KAN | Adam | CE | 0.974 | 0.874 |
| LTC | AdamW | LSCE | 0.918 | 0.493 |

When dealing with unbalanced data, accuracy alone is not enough, Macro-F1 shows how well you capture minority groups. The wide difference between Accuracy and Macro-F1 indicates that the model is missing minority groups, therefore Macro-F1 has been reported as the primary benchmark. Under the same conditions, Adam and CE, along with MLP and KAN, achieve high accuracy and macro-F1, while LTC, AdamW, and LSCE exhibit a large accuracy-macro-F1 gap, indicating poor minority-class performance. MLP and KAN demonstrate balanced performance with high accuracy and high Macro-F1.

For the MLP model, Adam/AdamW and CE yield the highest performance,

weighted Focal Loss configurations show significantly lower performance. For the KAN model, Adam/AdamW and CE/LSCE with KAN provide performance very close to MLP, SGD and weighted Focal Loss cause a significant decrease. The best performance for LTC was obtained using the AdamW optimizer and the Cross Entropy Loss function. With SGD and RMSProp, the performance of LTC was found to be approximately 82%. MLP and KAN achieved the highest success with Adam/AdamW optimization. Accuracy is greater than 97%, while Macro F1 is approximately 0.87–0.88. Focal Loss, while theoretically suitable for unbalanced classes, produced very low accuracy and F1 scores in practice. SGD proved insufficient in complex models, Adam and AdamW yielded better results.

Provided in the appendix, Table 2 shows the accuracy and Macro F1 results for MLP, KAN, and LTC.

## 6 Conclusion

This study compares different deep learning architectures, MLP, KAN, LTC, and various loss functions and optimization techniques on the MITBIH ECG dataset. Experimental results show that model performance is directly related to optimizer selection and loss function. MLP and KAN architectures provided the highest accuracy and balanced class performance when using Adam/AdamW optimization along with CE and LSCE losses. In contrast, Focal Loss configurations were insufficient for this dataset, and learning performance decreased significantly due to class imbalance. Although the LTC model offers theoretical advantages for time series modeling thanks to its continuous-time structure, it did not perform as well as MLP and KAN in practice.

These findings reveal that simple and powerful optimization strategies, such as Adam/AdamW and classical loss functions CE and LSCE, remain the most reliable choice for unbalanced health data. Future research suggests investigating methods such as class-weighted losses, sampling techniques, and hyperparameter optimization for Focal Loss to more effectively manage class imbalance. Furthermore, more advanced regularization and optimization strategies can be evaluated to improve the performance of continuous-time architectures like LTC.

Experimental analyses revealed that for the problem set under consideration, configuring the MLP or KAN architectures with the AdamW optimization algorithm and Label Smoothing Cross Entropy and CE loss functions yielded the most optimal result, approximately 0.98. The LTC architecture, however, requires a more selective hyperparameter optimization.

**Table 2.** Accuracy and Macro-F1 results for MLP, KAN, and LTC.

| Model | Loss | Optimizer | Test Accuracy | Macro F1 |
|---|---|---|---|---|
| MLP | CE | Adagrad | 0.958 | 0.723 |
| MLP | CE | Adam | 0.978 | 0.879 |
| MLP | CE | Adamw | 0.978 | 0.878 |

| | | | | |
|---|---|---|---|---|
| MLP | CE | RMSProp | 0.969 | 0.811 |
| MLP | CE | SGD | 0.912 | 0.484 |
| MLP | Focal_w | Adagrad | 0.148 | 0.285 |
| MLP | Focal_w | Adam | 0.155 | 0.288 |
| MLP | Focal_w | Adamw | 0.156 | 0.286 |
| MLP | Focal_w | RMSProp | 0.141 | 0.304 |
| MLP | Focal_w | SGD | 0.164 | 0.166 |
| MLP | LSCE | Adagrad | 0.957 | 0.730 |
| MLP | LSCE | Adam | 0.978 | 0.876 |
| MLP | LSCE | AdamW | 0.978 | 0.873 |
| MLP | LSCE | RMSProp | 0.967 | 0.825 |
| MLP | LSCE | SGD | 0.909 | 0.475 |
| KAN | CE | Adagrad | 0.956 | 0.684 |
| KAN | CE | Adam | 0.974 | 0.874 |
| KAN | CE | Adamw | 0.973 | 0.874 |
| KAN | CE | RMSProp | 0.944 | 0.696 |
| KAN | CE | SGD | 0.872 | 0.350 |
| KAN | Focal_w | Adagrad | 0.151 | 0.284 |
| KAN | Focal_w | Adam | 0.161 | 0.270 |
| KAN | Focal_w | Adamw | 0.230 | 0.296 |
| KAN | Focal_w | RMSProp | 0.130 | 0.210 |
| KAN | Focal_w | SGD | 0.027 | 0.019 |
| KAN | LSCE | Adagrad | 0.958 | 0.697 |
| KAN | LSCE | Adam | 0.973 | 0.865 |
| KAN | LSCE | Adamw | 0.974 | 0.867 |
| KAN | LSCE | RMSProp | 0.943 | 0.660 |
| KAN | LSCE | SGD | 0.866 | 0.331 |
| LTC | CE | Adagrad | 0.828 | 0.181 |
| LTC | CE | Adam | 0.908 | 0.466 |
| LTC | CE | Adamw | 0.914 | 0.484 |
| LTC | CE | RMSProp | 0.828 | 0.181 |
| LTC | CE | SGD | 0.828 | 0.181 |
| LTC | Focal_w | Adagrad | 0.025 | 0.009 |
| LTC | Focal_w | Adam | 0.025 | 0.009 |
| LTC | Focal_w | Adamw | 0.135 | 0.221 |
| LTC | Focal_w | RMSProp | 0.025 | 0.009 |
| LTC | Focal_w | SGD | 0.278 | 0.102 |
| LTC | LSCE | Adagrad | 0.828 | 0.181 |
| LTC | LSCE | Adam | 0.911 | 0.473 |
| LTC | LSCE | Adamw | 0.918 | 0.493 |
| LTC | LSCE | RMSProp | 0.828 | 0.181 |
| LTC | LSCE | SGD | 0.828 | 0.181 |

## References


1. Verpoort, P., Lee, A., & Wales, D. (2020). Archetypal landscapes for deep neural networks. Proceedings of the National Academy of Sciences, 117(36), 21857-21864. https://doi.org/10.1073/pnas.1919995117
2. Geiger, M., Spigler, S., d'Ascoli, S., Sagun, L., Baity-Jesi, M., Biroli, G. & Wyart, M. (2019). Jamming transition as a paradigm to understand the loss landscape of deep neural networks. Physical Review E, 100(1). https://doi.org/10.1103/physreve.100.012115
3. Baity-Jesi, M., Sagun, L., Geiger, M., Spigler, S., Arous, G., Cammarota, C., … & Biroli, G. (2019). Comparing dynamics: deep neural networks versus glassy systems. Journal of Statistical Mechanics Theory and Experiment, 2019(12), 124013. https://doi.org/10.1088/1742-5468/ab3281
4. Mehta, D., Zhao, X., Bernal, E., & Wales, D. (2018). Loss surface of XOR artificial neural networks. Physical Review E, 97(5). https://doi.org/10.1103/physreve.97.052307
5. Chaudhari, P., Choromanska, A., Soatto, S., LeCun, Y., Baldassi, C., Borgs, C., … & Zecchina, R. (2019). Entropy-SGD: biasing gradient descent into wide valleys*. Journal of Statistical Mechanics Theory and Experiment, 2019(12), 124018. https://doi.org/10.1088/1742-5468/ab39d9
6. Güler, İ. and Übeylı, E. (2005). An expert system for detection of electrocardiographic changes in patients with partial epilepsy using wavelet-based neural networks. Expert Systems,22(2),6271. https://doi.org/10.1111/j.1468-0394.2005.00295.x
7. Özesmi, S., Tan, C., & Özesmi, U. (2006). Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecological Modelling, 195(1-2), 83-93. https://doi.org/10.1016/j.ecolmodel.2005.11.012
8. Lopes, M., Rocha, B., Vieira, A., Sá, J., Rolim, P., & Silva, A. (2019). Artificial neural networks approaches for predicting the potential for hydropower generation: a case study for Amazon region. Journal of Intelligent & Fuzzy Systems, 36(6), 5757-5772. https://doi.org/10.3233/jifs-181604
9. Ahmad, M., Kumar, N., & Singh, B. (2016). Fast multilayer perceptron neural network-based control algorithm for shunt compensator in distribution systems. Iet Generation Transmission & Distribution, 10(15), 3824-3833. https://doi.org/10.1049/iet-gtd.2016.0328
10. Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. Ieee Access, 9, 4843-4873. https://doi.org/10.1109/access.2020.3048415
11. Sladojević, S., Arsenović, M., Anderla, A., Ćulibrk, D., & Stefanović, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience, 2016, 1-11. https://doi.org/10.1155/2016/3289801
12. Varjovi, M., Rahmanpour, M., Khosravi, M., Majdi, A., & Le, B. (2024). Performance Evaluation of Artificial Neural Networks and Support Vector Regression in Tunneling-Induced Settlement Prediction Incorporating Umbrella Arch Method Characteristics. International Journal of Engineering, 37(8), 1510-1521. https://doi.org/10.5829/ije.2024.37.08b.05
13. Rabiei, S., Jalilvand, E., & Tajrishy, M. (2021). A Method to Estimate Surface Soil Moisture and Map the Irrigated Cropland Area Using Sentinel-1 and Sentinel-2 Data. Sustainability, 13(20), 11355. https://doi.org/10.3390/su132011355

14. Patuwo, E., Hu, M., & Hung, M. (1993). Two-Group Classification Using Neural Networks*. Decision Sciences, 24(4), 825-845. https://doi.org/10.1111/j.1540-5915.1993.tb00491.x
15. Bousqaoui, H., Slimani, I., & Achchab, S. (2021). Comparative analysis of short-term demand predicting models using ARIMA and deep learning. International Journal of Electrical and Computer Engineering (Ijece), 11(4), 3319. https://doi.org/10.11591/ijece.v11i4.pp3319-3328
16. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539
17. Fei, R., Yao, Q., Zhu, Y., Xu, Q., Li, A., Wu, H., … & Hu, B. (2020). Deep Learning Structure for Cross-Domain Sentiment Classification Based on Improved Cross Entropy and Weight. Scientific Programming, 2020, 1-20. https://doi.org/10.1155/2020/3810261
18. Huang, Q., Le, J., Joshi, S., Mendes, J., Adluru, G., & DiBella, E. (2024). Arterial Input Function (AIF) Correction Using AIF Plus Tissue Inputs with a Bi-LSTM Network. Tomography, 10(5), 660-673. https://doi.org/10.3390/tomography10050051
19. Huang, J., Niu, G., Guan, H., & Song, S. (2023). Ultra-Short-Term Wind Power Prediction Based on LSTM with Loss Shrinkage Adam. Energies, 16(9), 3789. https://doi.org/10.3390/en16093789
20. Li, Y., Ren, X., Zhao, F., & Yang, S. (2021). A Zeroth-Order Adaptive Learning Rate Method to Reduce Cost of Hyperparameter Tuning for Deep Learning. Applied Sciences, 11(21), 10184. https://doi.org/10.3390/app112110184
21. Jia, P., Liu, H., Wang, S., & Wang, P. (2020). Research on a Mine Gas Concentration Forecasting Model Based on a GRU Network. Ieee Access, 8, 38023-38031. https://doi.org/10.1109/access.2020.2975257
22. Yu, X. (2023). The influence of regional tourism economy development on carbon neutrality for environmental protection using improved recurrent neural network. Frontiers in Ecology and Evolution, 11. https://doi.org/10.3389/fevo.2023.1146887
23. Wunderlich, T. and Pehle, C. (2021). Event-based backpropagation can compute exact gradients for spiking neural networks. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-91786-z
24. Yin, B., Corradi, F., & Bohté, S. (2023). Accurate online training of dynamical spiking neural networks through Forward Propagation Through Time. Nature Machine Intelligence, 5(5), 518-527. https://doi.org/10.1038/s42256-023-00650-4
25. Liao, X., Sahran, S., Abdullah, A., & Shukor, S. (2022). AdaCB: An Adaptive Gradient Method with Convergence Range Bound of Learning Rate. Applied Sciences, 12(18), 9389. https://doi.org/10.3390/app12189389
26. Tian, Y., Zhang, Y., & Zhang, H. (2023). Recent Advances in Stochastic Gradient Descent in Deep Learning. Mathematics, 11(3), 682. https://doi.org/10.3390/math11030682
27. Wang, Y., Zhou, P., & Zhong, W. (2018). An Optimization Strategy Based on Hybrid Algorithm of Adam and SGD. Matec Web of Conferences, 232, 03007. https://doi.org/10.1051/matecconf/201823203007
28. Chaudhury, S. and Yamasaki, T. (2021). Robustness of Adaptive Neural Network Optimization Under Training Noise. Ieee Access, 9, 37039-37053. https://doi.org/10.1109/access.2021.3062990
29. Albayati, A., Altaie, S., Al-Obaydy, W., & Alkhalid, F. (2024). Performance analysis of optimization algorithms for convolutional neural network-based handwritten digit recognition. Iaes International Journal of Artificial Intelligence (Ij-Ai), 13(1), 563. https://doi.org/10.11591/ijai.v13.i1.pp563-571

30. Liang, D., Ma, F., & Li, W. (2020). New Gradient-Weighted Adaptive Gradient Methods with Dynamic Constraints. Ieee Access, 8, 110929-110942. https://doi.org/10.1109/access.2020.3002590
31. Barkhordari, M. and Massone, L. (2022). Failure Mode Detection of Reinforced Concrete Shear Walls Using Ensemble Deep Neural Networks. International Journal of Concrete Structures and Materials, 16(1). https://doi.org/10.1186/s40069-022-00522-y
32. He, Y., Li, W., Zhang, W., Zhang, S., Pi, X., & Liu, H. (2021). Research on Segmentation and Classification of Heart Sound Signals Based on Deep Learning. Applied Sciences, 11(2), 651. https://doi.org/10.3390/app11020651
33. M.G, S. and Venkatesan, C. (2025). SwinDFU-Net: Deep learning transformer network for infection identification in diabetic foot ulcer. Technology and Health Care, 33(1), 601-618. https://doi.org/10.3233/thc-241444
34. Huang, T., Yang, R., Shen, L., Feng, A., Li, L., He, N., … & Lyu, J. (2022). Deep transfer learning to quantify pleural effusion severity in chest X-rays. BMC Medical Imaging, 22(1). https://doi.org/10.1186/s12880-022-00827-0
35. J. Xu and X. Fang, "ATT-BLKAN: A Hybrid Deep Learning Model Combining Attention is Used to Enhance Business Process Prediction," in IEEE Access, vol. 13, pp. 36175-36189, 2025, https://doi.org/10.1109/ACCESS.2025.3545071.

# A Dynamic Programming Approach for the Fuzzy Minimum Cost Maximum Flow Problem in the Multi-Stage Networks

Nermin Kartli
Ankara University – Department of Computer Science
E-posta: nkartli@ankara.edu.tr

**Abstract**

In this study, we consider the minimum-cost maximum-flow problem in a fuzzy environment for the multi-stage networks. In the problem under investigation, we model edge costs on the network with parametric fuzzy numbers. We transform uncertainty into deterministic subproblems using the $\alpha$-cut approach and propose a novel dynamic-programming-based heuristic algorithm.

In the proposed method, a dynamic-programming-based shortest-path algorithm, developed to accommodate the multi-stage network structure, is applied to the deterministic problem obtained for each $\alpha$ level. In this process, performed on the residual network, the capacity dominance criterion is considered alongside cost minimization, and paths with higher flow capacity are preferred among alternative paths with the same cost. This approach ensures a more balanced and stable solution process.

The algorithm updates the residual network at each increment step and defines reverse edges with negative costs. This results in an iterative improvement process similar to classical minimum cost flow methods. Left and right end solutions are calculated separately, the results are combined, and defuzzification is applied in accordance with fuzzy decision theory.

The effectiveness of the proposed algorithm was evaluated through numerical experiments performed on randomly generated multi-stage test networks. The results show that the developed method can produce consistent, stable, and computationally efficient solutions under different uncertainty levels. In this respect, the study offers a practical and applicable approach to solving fuzzy network flow problems.

**Keywords:** Fuzzy optimization, maximum flow at minimum cost, multi-stage networks, dynamic programming, alpha-cuts method, parametric fuzzy numbers, supply chain networks, heuristic algorithms.

## 1 Introduction

Network-based optimization problems are widely used in many fields such as logistics, transportation, production planning, supply chain management, telecommunications, and energy systems. Among these problems, the minimum cost maximum flow problem stands out as a fundamental model aiming to send the highest possible flow from a given source to a given destination at the lowest cost. In the classical case, this problem can be efficiently solved using linear programming techniques and specialized network algorithms.

However, in real-life applications, network parameters are often not precisely known and contain various uncertainties. Many external factors, such as transportation costs, fuel prices, exchange rates, demand fluctuations, weather conditions, and operational risks, cause system parameters to change over time. Classical models that do not consider such uncertainties may not adequately reflect the actual system behavior in practical applications and can produce misleading results for decision-makers.

Fuzzy set theory is widely used in modeling optimization problems defined under uncertainty. Fuzzy numbers allow for the representation of uncertain parameters without the need for probability distributions. Therefore, numerous studies have been conducted on network flow problems defined in a fuzzy environment in recent years. However, a significant portion of existing studies offer solutions based on classical shortest path algorithms on general network structures and do not adequately utilize the structural features of the problem.

Especially in layered structures such as multi-stage supply chain networks, classical shortest path algorithms lead to unnecessary computations and significantly increase the solution time. In such networks, nodes are divided into specific stages, and flow occurs only between successive stages. The effective use of this structure in the solution process is of great importance in terms of computational efficiency.

Although various alpha-cut approaches have been developed in the literature for fuzzy minimum cost maximum flow problems, most of these methods are not specifically designed for multi-stage network structures. Furthermore, capacity differences between alternative paths with the same cost are generally ignored in existing methods, which can negatively affect the total flow volume.

In this study, a new dynamic programming-based heuristic algorithm specific to multi-stage networks is proposed for solving the fuzzy minimum cost maximum flow problem. In the proposed method, fuzzy edge costs are modeled using a parametric representation and transformed into deterministic subproblems using the $\alpha$-cut approach. For each $\alpha$ level, a special shortest path algorithm based on dynamic programming, utilizing the hierarchical structure of the network, has been developed.

Furthermore, by applying a capacity dominance criterion among alternative paths with the same cost, paths with higher flow potential are preferred. This aims to minimize the total cost and increase the amount of flow sent. The solution process is systematically improved by iteratively updating the network structure.

The main contributions of this study can be summarized as follows:

- A new dynamic programming-based solution method is proposed for the fuzzy minimum cost maximum flow problem defined in multi-stage networks.
- Parametric fuzzy costs are efficiently solved by transforming them into deterministic problems using the $\alpha$-cut approach.
- More balanced and stable augmenting paths are obtained by using the capacity dominance criterion. The effectiveness of the proposed algorithm has been demonstrated through numerical experiments performed on randomly generated test networks.

The rest of the article is structured as follows: In the second section, we provide a summary of previous studies related to this study. The third section describes the problem in detail. The fourth section presents the proposed algorithm. The fifth section discusses the computational experiments and the results obtained. The final section provides general evaluations and suggestions for future research.

## 2 Related Work

Minimum cost flow problems are one of the fundamental problems that have long been studied in the field of classical network optimization. However, modeling cost and capacity parameters in networks with uncertainty makes the direct application of classical methods difficult. Studies on fuzzy minimum cost flow problems in the literature have addressed uncertainty models from different perspectives.

From a fundamental perspective, Mares and Horak (1983) [13] have defined fuzzy quantities in networks, and then Keyi and Wang [12] have considered the fuzzy minimum-cost flow problem

in a network. These studies are the first to address the effect of uncertain costs on network flow problems.

Ghatee and Hashemi have conducted research within the context of extended fuzzy flow models to address the fuzzy supply-demand balance and propose different solution methods. They have proposed the method within the scope of fuzzy multi-commodity flow, aiming to optimize the transportation of multiple products simultaneously under flow and cost uncertainty [10].

The study presented by Bozhenyuk and Gerasimenko (2013) [2] addresses the minimum cost flow problem in fuzzy dynamic networks and proposes a solution method based on the fuzzification of parameters such as sub-stream bounds, transit cost, and time. This study focuses on integrating fuzzy costs and transit times into the flow process.

Similarly, Bozhenyuk et al. [3] presented a method for determining minimum cost flow in fuzzy dynamic networks using substream bounds and costs. This study has developed approaches to flow problems in uncertain environments where dynamic parameters can change over time.

In the study by El-Sherbeny (2019) [8], a model that fuzzifies time windows as a valid extension of the classification problem is proposed, and an algorithm suitable for this model is developed. This study presents a path-based solution using array-by-array augmenting by combining fuzzy costs and time windows.

In the study [6] by Chung and Duan, multi-stage fuzzy neural network (MSFNN) models were developed to overcome the "curse of dimensionality" problem encountered by single-stage fuzzy neural networks in high-dimensional problems. The authors divided the fuzzy inference process into multiple stages by defining two hierarchical network structures: incremental and aggregated. This structure prevents the exponential increase of the number of fuzzy rules and provides a more efficient modeling approach. In addition, structural and parameter learning algorithms of the proposed models were developed, and their effectiveness was demonstrated through various numerical experiments. This work focuses on prediction and modeling problems by addressing multi-stage fuzzy inference within the framework of neural networks; it is not directly related to network flows or optimization-based decision problems. In contrast, our work develops a dynamic programming-based optimization algorithm for solving the minimum cost maximum flow problem in a fuzzy environment and focuses on the decision-making process on network structures.

Ding et al. [7] transform the target controllability problem with maximum covered nodes set in multiplex networks into a minimum-cost maximum-flow problem based on graph theory. They have proposed an algorithm called target minimum-cost maximum-flow (TMM).

In his study, Baidya [1] modeled the uncertainty in multi-stage supply chain networks using grey numbers and solved the problem by transforming it into deterministic sub-models. In the study, the objective function in interval form was made single-objective using the mean-width method, and the models were solved with the help of LINGO.

The studies [4, 5] propose a novel algorithmic framework capable of solving minimum-cost and maximum-cost flow problems in near linear time. The method is based on a combination of interior point methods and dynamic graph data structures, updating the flow incrementally over approximate minimum-rate cycles. This yields one of the fastest known theoretical solutions for very large-scale and deterministic networks. However, in this study, edge costs and capacities are treated as exact (deterministic), and uncertainty or fuzzy structures are not considered. Furthermore, the algorithm is designed for generalized directed networks and does not include multi-stage specialized network structures or $\alpha$-cut-based fuzzy solution approaches.

In the study [14], decision problems on multi-stage and uncertain network structures are addressed within the framework of fuzzy set theory. Edge costs and system parameters are modeled with fuzzy numbers, and uncertainties are transformed into deterministic subproblems using the $\alpha$-cut approach. The resulting subproblems are solved using classical network optimization methods, and the results are reduced to a single decision value using defuzzification techniques.

This study has conceptual similarities with the proposed method in that it offers a systematic framework for modeling and solving multi-stage networks in a fuzzy environment. While the aforementioned study focuses on modeling fuzzy parameters in multi-stage networks and creating a general solution framework, in this study, we propose a dynamic programming-based shortest path algorithm specifically developed for multi-stage network structures. Furthermore, while the current study uses general-purpose classical network solution methods, the proposed approach considers the network structure and capacity dominance criteria together, offering a more efficient solution mechanism specific to the fuzzy minimum cost maximum flow problem.

# 3 Problem Statement

In this section, we define the fuzzy minimum cost maximum flow problem on a multi-stage directed network.

## 3.1 Network Structure

We assume that a directed network $G = (V, E)$ is given, where $V$ represents the set of nodes and $E$ represents the directed edge set. We also assume that the considered network can be divided into $m + 1$ stages as follows:

$$V = V_0 \cup V_1 \cup \cdots \cup V_m \tag{1}$$

Here,

- $V_0 = \{S\}$ represents the source node,
- $V_m = \{T\}$ represents the destination node,
- $V_1, V_2, \ldots, V_{m-1}$ represent the intermediate stages

The network structure is defined such that there is only a connection between successive stages. Accordingly, the following condition is satisfied:

$$(i, j) \in E \Rightarrow i \in V_k, j \in V_{k+1}, \quad k = 0, 1, \ldots, m - 1 \tag{2}$$

## 3.2 Capacities and Flow Variables

A positive capacity value is defined for each $(i, j) \in E$ edge:

$$c_{ij} > 0 \tag{3}$$

The flow variable $x_{ij}$ represents the amount of flow sent through the $(i, j)$ edge and satisfies the following condition:

$$0 \leq x_{ij} \leq c_{ij}, \quad \forall (i, j) \in E \tag{4}$$

The flow conservation condition for the nodes is defined as follows:

$$\sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = \begin{cases} F, & i = S \\ -F, & i = T \\ 0, & \text{other cases} \end{cases} \tag{5}$$

Here, $F$ represents the total flow amount sent from node $S$ to node $T$.

### 3.3 Fuzzy Edge Costs

For each edge $(i, j) \in E$, the unit cost is defined as a parametric fuzzy number:

$$\tilde{w}_{ij} = (u_{ij}(r), v_{ij}(r)), \quad r \in [0, 1] \tag{6}$$

Here, the functions $u_{ij}(r)$ and $v_{ij}(r)$ satisfy the following properties:

- $u_{ij}(r)$ is a bounded, non-decreasing and left-continuous function,
- $v_{ij}(r)$ is a bounded, non-increasing and left-continuous function,
- $u_{ij}(r) \leq v_{ij}(r)$, $\forall r \in [0, 1]$.

This definition allows for the parametric modeling of cost uncertainty.

### 3.4 $\alpha$-cuts-Based Deterministic Model

Fuzzy costs are transformed into deterministic problems using the $\alpha$-cuts approach. For each $\alpha \in [0, 1]$, the edge costs are obtained as follows:

$$w_{ij}^{L}(\alpha) = u_{ij}(\alpha) \tag{7}$$

$$w_{ij}^{R}(\alpha) = v_{ij}(\alpha) \tag{8}$$

Accordingly, two separate deterministic problems are defined for each $\alpha$ level.

### 3.5 Objective Function

For each $\alpha$ level, the objective is to minimize the total cost among the solutions that provide the maximum flow rate for the left end side:

$$\min \sum_{(i,j)\in E} w_{ij}^{L}(\alpha) x_{ij} \tag{9}$$

$$\max F \tag{10}$$

Similarly, for the right end side, we solve the following problem:

$$\min \sum_{(i,j)\in E} w_{ij}^{R}(\alpha) x_{ij} \tag{11}$$

$$\max F \tag{12}$$

This objective function is considered in lexicographic order, first aiming for the minimum cost, then the maximum flow rate.

### 3.6 Definition of Fuzzy Solution

By combining the left and right end solutions obtained for each $\alpha$ level, the fuzzy minimum cost function is defined as follows:

$$\tilde{M}(r) = (M_L(r), M_R(r)), \quad r \in [0, 1] \tag{13}$$

Here, $M_L(r)$ and $M_R(r)$ represent the left and right end cost functions, respectively.

### 3.7 Assumptions

The following assumptions are accepted in this study:

- The mesh structure is multi-stage and does not contain cycles.
- Edge capacities are exact and positive.
- Edge costs are modeled with parametric fuzzy numbers.
- The flow is divisible.
- Each $\alpha$ level is solved independently.

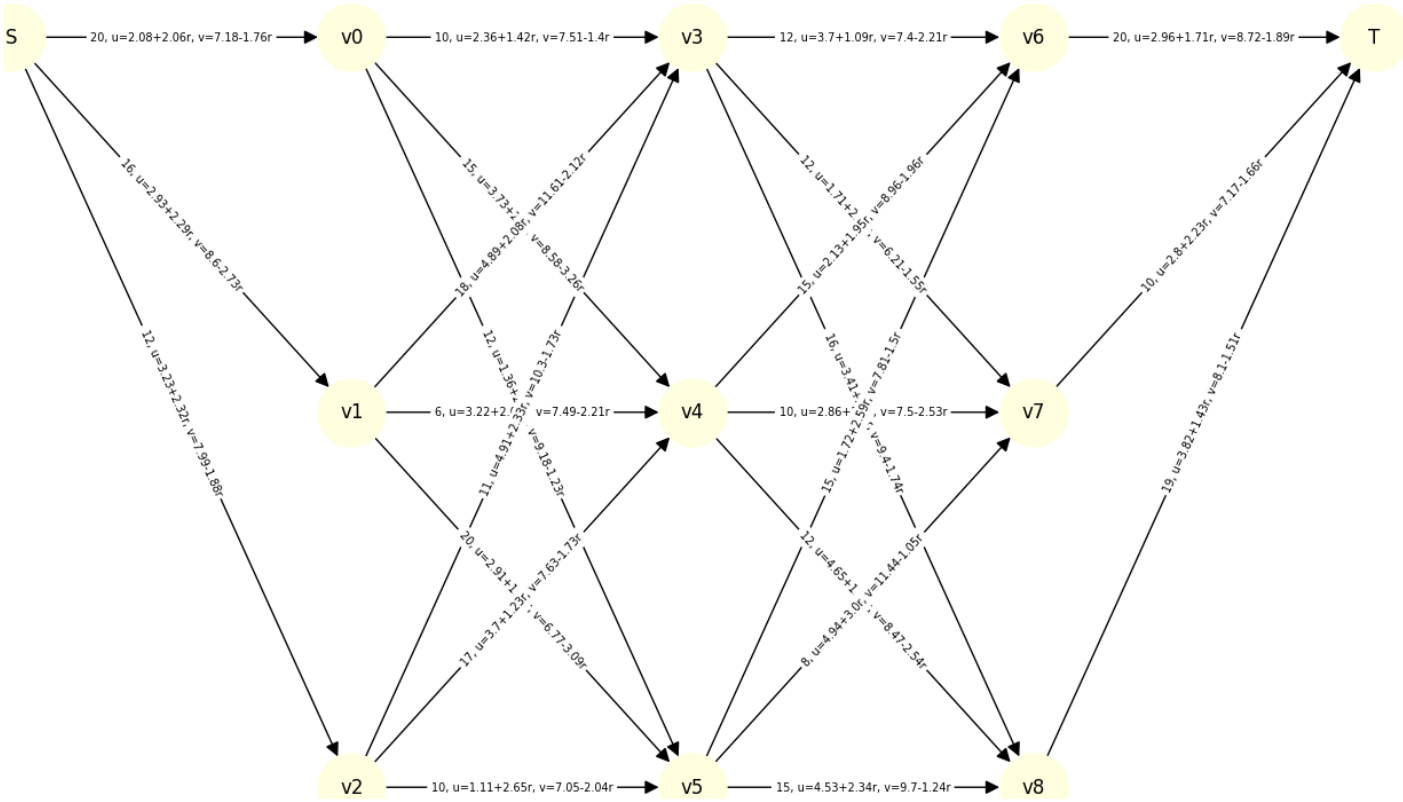



Figure 1: Fuzzy Minimum Cost Maximum Flow Problem:Example

## 4 Proposed Algorithm

We propose a new heuristic algorithm based on $\alpha$-cuts and dynamic programming to solve the fuzzy minimum cost maximum flow problem defined on multi-stage supply chain networks.

The proposed method uses the $\alpha$-cuts approach to reduce fuzzy costs to deterministic problems and solves a separate minimum cost maximum flow problem for each $\alpha$ level. The obtained solutions are then combined to form a fuzzy solution function. Thanks to this, fuzzy problems are transformed into numerous deterministic problems, as in the studies [9, 11].

The algorithm consists of two main components:

- $\alpha$-cuts based fuzzy solution loop,
- Dynamic programming-based shortest path algorithm specific to the multi-stage network structure.

### 4.1 $\alpha$-cuts Based Solution Approach

Fuzzy costs are defined in the form $(u_{ij}(r), v_{ij}(r))$ using a parametric representation. Here, $r \in [0, 1]$, where each value of $r$ represents a $\alpha$-cut level.

In the algorithm, the interval $[0, 1]$ is divided into $N$ parts with equal steps, and the $\alpha$ levels are obtained as follows:

$$r_k = \alpha_k = \frac{k}{N}, \quad k = 0, 1, \ldots, N \tag{14}$$

For each $\alpha$ value, the deterministic edge costs are calculated as follows:

- For the left end side, $w_{ij}^L(\alpha) = u_{ij}(\alpha)$,
- For the right end side, $w_{ij}^R(\alpha) = v_{ij}(\alpha)$.

In this way, a classic minimum cost maximum flow problem is obtained for each $\alpha$ level.

### 4.2 Residual Network and Flow Update Mechanism

For each $\alpha$ level, the residual network is initially defined as being the same as the original network.

The following steps are applied in each iteration:

1. The shortest path between $S$ and $T$ is calculated on the residual network.
2. The smallest capacity value on this path is determined as the amount of flow that can be increased.
3. The determined amount is sent along the path.
4. The capacities of the edges used are reduced.
5. The residual network is updated by adding reverse-directional edges.
6. The costs of reverse-directional edges are assigned as negative.

This process continues until there is no path left between $S$ and $T$ on the residual network.

### 4.3 Dynamic Programming Based Shortest Path Algorithm

The network is given in stages as follows:

$$V = V_0 \cup V_1 \cup \cdots \cup V_m \tag{15}$$

Here,

- $V_0 = \{S\}$,
- $V_m = \{T\}$,
- $V_1, \ldots, V_{m-1}$ represent intermediate stages.

If there are multiple paths with equal cost, the capacity value is considered.

Let $cost[u, v]$ denote the unit cost of flow on edge $(u, v)$, and let $c[u, v]$ denote the capacity of edge $(u, v)$. For example, for the left end side and an $\alpha$, $cost[u, v] = w_{ij}^L(\alpha) = u_{ij}(\alpha)$.

Let $M[i, v]$ represent the minimum unit cost of a path from the source node $S$ to node $v$ in stage $i$. Let $CA[i, v]$ denote the minimum capacity along the path corresponding to this minimum unit cost.

The initial conditions are defined as:

$$M[1, v] = cost[S, v], \quad CA[1, v] = c[S, v].$$

For each stage $i \geq 1$ and for each $v \in V_{i+1}$, the recursive relation is given by:

$$M[i+1, v] = \min_{u \in V_i} \{M[i, u] + cost[u, v]\}.$$

Let $u^*$ be the node at which the minimum is attained. If there are multiple paths with equal cost, the capacity value is considered. In this case, we choose a node with maximum flow. Then, the corresponding capacity is updated as:

$$CA[i+1, v] = \min \{CA[i, u^*],\ c[u^*, v]\}.$$

### 4.4 Capacity Dominance Criterion

In the proposed method, when choosing between alternative paths with the same cost, not only the cost but also the minimum capacity value on the path is considered.

For two paths:

$$M_1 = M_2 \tag{16}$$

If this is the case, the following criterion is applied:

$$\max\{\min(CA_1, c_1), \min(CA_2, c_2)\} \tag{17}$$

In this way, the path with higher capacity is preferred.

This approach allows the algorithm to produce more balanced solutions.

### 4.5 Creating the Fuzzy Solution

After obtaining the left and right end solutions separately, the cost pairs are combined for each $\alpha$ level to create a fuzzy minimum cost function:

$$\tilde{M}(r) = (M_L(r), M_R(r)) \tag{18}$$

Here, $M_L(r)$ and $M_R(r)$ represent the left and right end cost functions, respectively.

If desired, the fuzzy solution can be reduced to a single deterministic value using defuzzification methods.

In this study, the weighted center method was used.

### 4.6 Advantages of the Algorithm

The main advantages of the proposed method are summarized below:

- It is specifically designed for multi-stage mesh structures.
- It has lower computational costs compared to the classic Bellman-Ford algorithm.
- It produces more stable solutions thanks to capacity dominance.
- It can directly take fuzzy costs into account.
- It has high applicability to large-scale problems.

**Algorithm 1:** Fuzzy Minimum-Cost Maximum-Flow Algorithm

**Input :** Directed multi-stage graph $G = (V, E)$,
Source node $S$, Sink node $T$,
Edge capacities $c_{ij}$,
Fuzzy costs $\tilde{w}_{ij} = (u_{ij}(r), v_{ij}(r))$,
Number of $\alpha$-levels $N$

**Output:** Fuzzy minimum cost $\tilde{M}(r)$,
Maximum flow $F$

```
Initialize α_k = k/N, k = 0, 1, ..., N
Initialize result lists L, R
foreach mode ∈ {left, right} do
    Set residual graph R ← G
    Set total flow F ← 0
    Set total cost M ← 0
    foreach α ∈ α_k (increasing for left, decreasing for right) do
        Compute deterministic costs w_ij(α) from fuzzy costs
        Update edge weights in R using w_ij(α)
        while true do
            (P, f, m) ← DP_Shortest_Path(R, S, T)
            if P = ∅ then
                break
            Augment flow f on path P
            Update residual graph R
            Update F ← F + f
            Update M ← M + f · m
        Store (α, F, M) in L or R
Construct fuzzy cost function M̃(r) from L and R
return M̃(r), F
```

**Algorithm 2:** Dynamic Programming Shortest Path with Capacity Dominance

**Input :** Residual graph $R$,
Source $S$, Sink $T$,
Stage partition $\{V_0, V_1, \ldots, V_m\}$,
Edge weights $w_{ij}$, capacities $c_{ij}$

**Output:** Shortest path
Bottleneck capacity
Path cost

```
Initialize tables M[i][v] ← ∞, CA[i][v] ← 0, P[i][v] ← NULL
foreach v ∈ V_1 do
    if (S, v) ∈ E then
        M[1][v] ← w_Sv
        CA[1][v] ← c_Sv
        P[1][v] ← S
for i = 1 to m − 1 do
    foreach v ∈ V_{i+1} do
        best ← ∞
        bestCap ← 0
        foreach u ∈ V_i do
            if (u, v) ∈ E and M[i][u] < ∞ then
                val ← M[i][u] + w_uv
                cap ← min(CA[i][u], c_uv)
                if val < best then
                    best ← val
                    bestCap ← cap
                    P[i + 1][v] ← u
                else if val = best and cap > bestCap then
                    bestCap ← cap
                    P[i + 1][v] ← u
        M[i + 1][v] ← best
        CA[i + 1][v] ← bestCap
    if all M[i + 1][v] = ∞ then
        return ∅, 0, ∞
Reconstruct path p from table P
return p, CA[m][T], M[m][T]
```

### 4.7 Defuzzification Method

After obtaining the fuzzy minimum cost function $\tilde{M}(r) = (M_L(r), M_R(r))$ it needs to be converted into a single deterministic value that can be used in the decision-making process. For this purpose, the defuzzification process is applied.

In this study, the weighted centroid method, which is widely used in fuzzy decision theory, has been preferred. In this method, the fuzzy cost function is defuzzified as follows:

$$M^* = \frac{\int_0^1 \frac{M_L(r)+M_R(r)}{2} \cdot r \, dr}{\int_0^1 r \, dr} \tag{19}$$

If discrete $\alpha$ levels are used, this expression is approximately calculated in the following form:

$$C^* \approx \frac{\sum_{k=0}^{N} \frac{M_L(\alpha_k)+M_R(\alpha_k)}{2} \cdot \alpha_k}{\sum_{k=0}^{N} \alpha_k} \tag{20}$$

Here, $N$ represents the number of levels $\alpha$ used.

This method produces a decision value that represents the uncertainty structure in a balanced way by considering both the left and right endpoints of the fuzzy cost function. It is also preferred because of its computationally simple nature and ease of implementation.

# 5 Computational Experiments and Results

In this section, the effectiveness and performance of the proposed algorithm are evaluated through numerical experiments conducted on randomly generated multi-stage networks. All experiments were performed in a simulation environment developed using the Python programming language.

## 5.1 Experimental Setup

Test problems were randomly generated considering different network sizes and number of stages. In each test network, nodes were assigned to specific stages, and edges were created such that connections only exist between consecutive stages.

The basic parameters used in the experiments are as follows:

- Number of stages: $m = 3, 4, 5$
- Number of nodes in each stage: Between 3 and 10
- Edge capacities: Random integers in the range $[5, 50]$
- Edge costs: Parametric fuzzy numbers
- Number of $\alpha$ levels: $N = 10$

Fuzzy cost functions were constructed using linear and triangular parametric structures. To ensure the reproducibility of the experiments, random number generation was performed using a fixed initial value (seed).

All experiments were conducted on a personal computer with an Intel i7 processor and 16 GB of RAM.

## 5.2 Evaluation Criteria

The performance of the proposed method was evaluated using the following criteria:

- Total maximum flow rate sent,
- Total fuzzy cost value,
- Defuzzified cost value,
- Solution time,
- Stability under different $\alpha$ levels.

In addition, the obtained results were compared with classical minimum cost maximum flow algorithms commonly used in the literature.

### 5.3 Experimental Results

For each test problem, left and right end solutions were calculated separately, and the resulting fuzzy cost functions were created. The results obtained for an example test problem are presented in Table 1.

Table 1: Experimental Results for a Sample Test Problem

| $\alpha$ | Flow | Left End Cost | Right End Cost |
|---|---|---|---|
| 0.0 | 48 | 525.83 | 1114.47 |
| 0.2 | 48 | 607.16 | 1201.67 |
| 0.4 | 48 | 687.43 | 1278.98 |
| 0.6 | 48 | 768.86 | 1356.14 |
| 0.8 | 48 | 843.71 | 1433.03 |
| 1.0 | 48 | 923.74 | 1510.85 |

When the results are examined, it is observed that the maximum flow rate remains constant across different $\alpha$ levels, while the total cost varies with the level of uncertainty. This shows that the proposed algorithm produces stable and consistent solutions.

Figure 2 shows the left and right end cost curves obtained for different $\alpha$ levels.

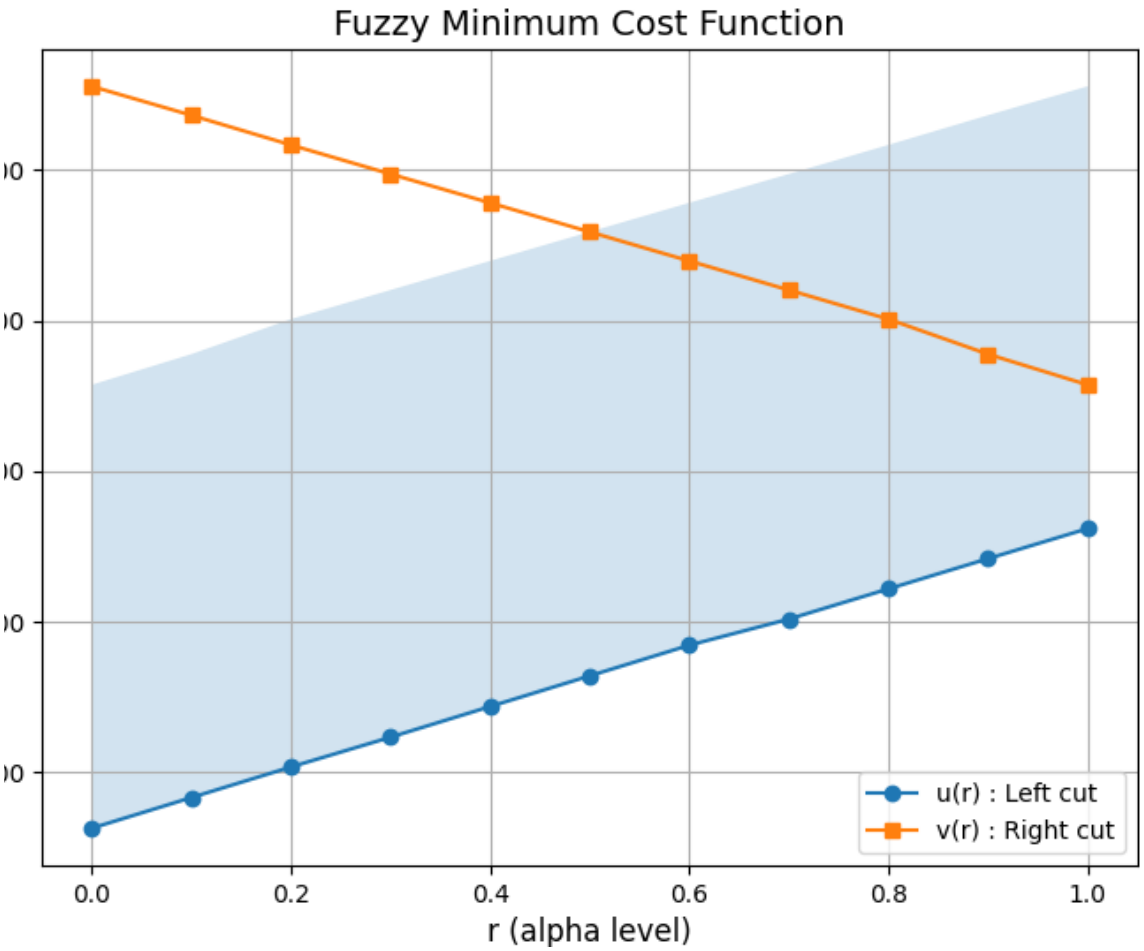



Figure 2: Variation of left and right end costs according to $\alpha$ levels

### 5.4 Comparative Analysis

The proposed method was compared with the classical Bellman-Ford-based minimum-cost flow algorithm. Experimental results show that the developed dynamic programming-based approach provides a lower solution time in multi-stage network structures.

It has been observed that the computational cost of classical methods increases rapidly, especially as the number of nodes increases, while the proposed method maintains its scalability.

The experimental findings reveal that the proposed algorithm produces stable solutions under different levels of uncertainty. Thanks to the capacity dominance criterion, selecting paths with higher flow capacity among paths with the same cost positively affects the overall performance.

Furthermore, the dynamic programming-based shortest path algorithm offers a significant computational advantage compared to classical approaches by enabling the efficient use of the multi-stage network structure.

## 6 Conclusions and Future Work

In this study, a new dynamic programming-based heuristic algorithm is proposed for solving the fuzzy minimum cost maximum flow problem defined on multi-stage networks. The proposed method transforms parametric fuzzy costs into deterministic subproblems using the $\alpha$-section approach and produces efficient solutions by taking advantage of the staged structure of the network.

The developed algorithm provides lower computational costs compared to classical shortest path-based methods and obtains more balanced growth paths thanks to the capacity dominance criterion. Computational experiments have shown that the method produces stable, consistent, and scalable solutions under different levels of uncertainty.

The results indicate that the proposed approach can be effectively applied, especially in layered structures such as multi-stage supply chain networks. Furthermore, parametric modeling of fuzzy costs ensures a more realistic reflection of uncertainty in the decision-making process.

Future studies plan to improve the proposed method in the following areas:

- Integration of network models containing fixed charges into the algorithm,
- Modeling capacity values in a fuzzy manner,
- Addition of environmental and temporal criteria within the framework of multi-objective optimization,
- Implementation of applications on large-scale real datasets,
- Further reduction of solution time using parallel and distributed computing techniques.

Accordingly, it is believed that the proposed approach will form an important foundation for future studies in the field of fuzzy network optimization.

## References

[1] Abhijit Baidya. Application of grey number to solve multi-stage supply chain networking model. *International Journal of Logistics Systems and Management*, 47(4):494–518, 2024.

[2] Alexander Bozhenyuk and Evgeniya Gerasimenko. Algorithm for monitoring minimum cost in fuzzy dynamic networks. *Information Technology and Management Science*, 16(1):53–59, 2013.

[3] Alexander Bozhenyuk, Evgeniya Gerasimenko, Igor Rozenberg, and Irina Perfilieva. Method for the minimum cost maximum flow determining in fuzzy dynamic network with nonzero flow bounds. In *Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology*, pages 385–392. Atlantis Press, 2015.

[4] Li Chen, Rasmus Kyng, Yang Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. *Journal of the ACM*, 72(3):1–103, 2025.

[5] Li Chen, Rasmus Kyng, Yang P Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Almost-linear-time algorithms for maximum flow and minimum-cost flow. *Communications of the ACM*, 66(12):85–92, 2023.

[6] Fu-Lai Chung and Ji-Cheng Duan. On multistage fuzzy neural network modeling. *IEEE Transactions on fuzzy systems*, 8(2):125–142, 2000.

[7] Jie Ding, Changyun Wen, Guoqi Li, Pengfei Tu, Dongxu Ji, Ying Zou, and Jiangshuai Huang. Target controllability in multilayer networks via minimum-cost maximum-flow method. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1949–1962, 2021.

[8] Nasser A. El-Sherbeny. The fuzzy minimum cost flow problem with the fuzzy time-windows. *American Journal of Applied Mathematics and Statistics*, 7(6):191–195, 2019.

[9] Nizami A Gasilov and Şahin Emrah Amrahov. A survey and comparative analysis of different approaches to fuzzy differential equations modeling dynamics with uncertain parameters of deterministic character. *Kybernetika*, 61(3):289–347, 2025.

[10] Mehdi Ghatee and S Mehdi Hashemi. Some concepts of the fuzzy multicommodity flow problem and their application in fuzzy network design. *Mathematical and Computer Modelling*, 49(5–6):1030–1043, 2009.

[11] Nermin Kartli, Erkan Bostanci, and Mehmet Serdar Guzel. Heuristic algorithm for an optimal solution of fully fuzzy transportation problem. *Computing*, 106(10):3195–3227, 2024.

[12] Wang Keyi and Wang Zhongtuo. Fuzzy minimum-cost flow in network and its application in transportation problems. In *Control Science and Technology for Development*, pages 407–410. Elsevier, 1986.

[13] Milan Mares and Josef Horák. Fuzzy quantities in networks. *Fuzzy Sets and Systems*, 10(1-3):123–134, 1983.

[14] Xiaodong Shi, Xiangping Zhai, Rui Wang, Yi Le, Shuang Fu, and Ningzhong Liu. Task planning of multiple unmanned aerial vehicles based on minimum cost and maximum flow. *Sensors*, 25(5):1605, 2025.

# Extraction of Features Characterizing Keystroke Dynamics in the User Identification problem

Makhamadaziz Rasulmukhamedov[1][0000-0001-8404-3013], Gulmira Mirzaeva[2][0000-0002-0412-3951], Nomaz Mirzaev[2][0000-0002-0657-2654], Nuraddin Gafforov[1,3][0000-0003-4905-3675]

[1]Tashkent State Transport university, Tashkent, Republic of Uzbekistan
[2]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Republic of Uzbekistan
[3]Digital technologies and artificial intelligence research institute, Tashkent, Republic of Uzbekistan
mrgaffarov28@gmail.com

**Abstract.** This article examines the separation of characteristics describing the user's behavior in the matter of identifying the user based on keyboard dynamics. The study analyzed time sequences recorded by users while working with the keyboard. Haar wavelet transformation was used to distinguish characteristics that characterize user properties. The obtained wavelet coefficients were used as characters characterizing users. The advantages of the highlighted features were checked using SVM, MLP, Random Forest, and KNN algorithms. The results obtained can be used in the development of biometric authentication systems.

**Keywords:** Keyboard dynamics, user identification, Haar wavelet transform, SVM, MLP, KNN, Random Forest, time series, hold time, flight time.

## 1 Introduction

In the modern digital technology environment, ensuring information security remains one of the pressing problems. In particular, methods that ensure reliable user authentication are an integral part of any information system. Traditional authentication systems based on passwords, PIN codes, and physical tokens, despite their ease of use, have serious shortcomings in terms of information security. In addition, the high probability of these tools being forgotten under the influence of subjective factors, appropriated by third parties, as well as detected or compromised as a whole as a result of various cyberattacks reduces their reliability [1]. Biometric authentication technologies, including fingerprint or facial recognition, voice identification methods, despite their high accuracy, are not widely used in all fields due to the need for special hardware and software tools for their implementation, as well as the existence of possible risks to the confidentiality of personal biometric data [2].

Therefore, in recent years, passive biometric systems based on tracking user-device interaction have attracted special attention. In particular, dynamic authentication methods based on the user's keyboard typing actions are considered one of the promising directions in this direction. Each user's typing style when working with the keyboard, in particular the interval between pressing and releasing keys, the time intervals between keystrokes, typing speed, and rhythmic characteristics, are individual, and it is possible to identify a person based on these characteristics [3, 4].

This study proposes the use of wavelet transformation to implement authentication based on user keyboard typing actions. Wavelet transform allows for the identification of complex signal features through time-frequency analysis, and in the process serves to extract unique user-specific parameters [5]. This approach is especially effective in determining the temporal and local characteristics of user activity.

## 2 Methodology

### 2.1 General Description of the Study

This research aims to solve the problem of biometric authentication based on keyboard dynamics and aims to deeply analyze typing time signals in the user identification process. The main focus of the research is on identifying the time-varying and uneven properties of data entered through the keyboard, as well as identifying biometric features specific to the user from these characteristics [7].

The research process includes the following sequential steps: using a ready-made keyboard dynamics dataset, pre-processing the raw data, feature extraction based on wavelet, user identification using machine learning algorithms, and evaluating the final results based on statistical criteria. These stages are interconnected and serve to ensure the reliability and practical effectiveness of the proposed methodology.

The research results create a scientific basis for further improvement of authentication systems based on keyboard dynamics and their implementation in real information security systems [8].

### 2.2 Dataset

To study the problem of identifying users based on keyboard dynamics, an open and widely used dataset called “DSL-StrongPasswordData.csv” was used [3, 6]. This dataset was presented by K.S. Killourhy and R.A. Maxion at the DSN-2009 conference. A total of 51 users participated in the dataset, with each user entering the same predetermined text sequence (similar to a static password) several hundred times [4]. As a result, there are an average of 400 writing samples for each user, which allows for the study of stable statistical characteristics of individual writing style [9,16].

Each record in the dataset is based on key press and key release timestamps. The total data volume is over 20,400 records. Each record contains over 30 time-related

features, mainly of the Hold Time (Dwell Time) and Flight Time types. Hold Time represents the difference between the time a button is pressed and released, while Flight Time describes the time interval between two consecutive button presses. These time parameters incorporate the speed, rhythm, and motor characteristics specific to individual users [3,4].

One of the important advantages of the dataset is that it is open source and allows for direct comparison of research results with other scientific works. Therefore, using this dataset allows for an objective assessment of the reliability and effectiveness of the proposed wavelet-based approach.

### 2.3 Data Preprocessing

In the study, data describing keyboard typing behavior was expressed as a time-dependent feature vector for each user [10]. Each writing sample is represented by a set of properties of the following type:

$$X = \{x_1, ..., x_i, ..., x_n\} \quad (1)$$

here $x_i$– is one of the time parameters recorded during keyboard operation, $n$– is the number of common features allocated for one recording sample.

**Determining incorrect and extreme values.** Extreme values arising as a result of random pauses or technical delays that may occur in the set of properties can negatively affect the model's efficiency. Therefore, each element in vector $X$ was passed through a statistical filter:

$$x_i \in \left[\mu_x - 3\sigma_x,\, \mu_x + 3\sigma_x\right]$$

Values outside this interval were assessed as noise, removed, or replaced with threshold values [11].

**Normalize features.** In order to reduce the overall typing speed differences between users and bring all features to a single scale, Z-score normalization was applied. As a result, the normalized feature vector is expressed as:

$$x_i^{norm} = \frac{x_i - \mu_x}{\sigma_x}$$

As a result of this process, all features were centered around zero and had a variance of one, which ensures stable and accurate operation of machine learning algorithms.

**Standardization of time series.** To apply the wavelet transform, all $X$ vectors were scaled to the same size. To do this, short sequences were zero-padded, and long sequences were truncated to a specified maximum length:

$$x_i \in \mathrm{R}$$

As a result, the pre-processed feature set was made ready for wavelet-based multi-stage analysis [14,15].

### 2.4 Wavelet-Based Feature Extraction

Since time series related to keyboard dynamics are highly variable and have local characteristics, it is advisable to use a multi-resolution approach in their analysis [17]. Therefore, the discrete Haar wavelet transform was used in this study. The Haar wavelet has a simple structure, is effective in detecting sharp signal changes, and allows identifying time symbols specific to the keyboard recording process [18].

Let's isolate $N$ sequentially from the previously given data (1), which can be represented as:

$$\mathfrak{X} = (x_1, ..., x_i, ..., x_N), \tag{2}$$

Here $x_i$ – elements consisting of the values of Hold Time and Flight Time recorded during keyboard recording ($i = \overline{1,N};\ N = 2^n$).

It should be noted that the sequence of elements of the vector $\mathfrak{X}$ consists of discrete values, measured in milliseconds. In order to apply the Haar wavelet transform, it was ensured that $N$ has a degree of 2.

The Haar wavelet transform processes the signal based on pairwise elements. Each pair is decomposed using the following two orthogonal basis functions

$$\varphi = \left(\frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}}\right),\ \psi = \left(\frac{1}{\sqrt{2}}; -\frac{1}{\sqrt{2}}\right)$$

We will group the selected (2) initial signal sequence in pairs:

$$(x_1, x_2), (x_3, x_4), ..., (x_{N-1}, x_N) \tag{3}$$

The approximation and detail coefficients are calculated for each of the separated (3) pairs:

$$C_{1,k} = \frac{x_{2k-1} + x_{2k}}{\sqrt{2}}, \tag{4}$$

$$d_{1,k} = \frac{x_{2k-1} - x_{2k}}{\sqrt{2}}, \tag{5}$$

$$k = 1,2..., [N/2].$$

Here:

$C_{1,k}$— represents the low-frequency (general) component of the signal,

$d_{1,k}$— represents the high-frequency (rapidly changing) component.

The obtained approximation coefficients are decomposed again using the Haar wavelet in the next step. Thus, using formulas (4) and (5) sequentially, we generate the following vector:

$$\left(C_J, d_J, d_{J-1}, ..., d_1\right),$$

Here $J = \log_2 N$ – maximum decomposition rate. As a result, all the coefficients obtained fully represent the original signal and incorporate features at different time scales.

In this study, the approximation and detail coefficients obtained as a result of the Haar wavelet transform were used directly as a set of features, without performing additional statistical aggregation [19,20]. That is, the feature vector for each writing sample was formed as follows:

$$F = [C_J, d_J, d_{J-1}, ..., d_1]$$

This approach allows for maximum preservation of the time-scale structure of keyboard dynamics and serves to transfer user-specific typing patterns to classification algorithms without losing them [19,21].

### 2.5 User identification algorithms

Several machine learning algorithms were tested to identify users based on the extracted features. Including:

- Support Vector Machine (SVM);
- K-Nearest Neighbors, (KNN);
- Random Forest (RF);
- Multilayer Perceptron (MLP);

Each model was trained using 70% of the dataset and tested on the remaining 30%. This distribution was applied to evaluate the models' generalization capability and to ensure the objectivity of the results.

## 3 Results

### 3.1 Analysis of extracted features using Haar wavelet transform

The features obtained based on the Haar wavelet transform made it possible to reflect the time structure of the keyboard typing process at different scales. The zoom and detail coefficients obtained as a result of multi-step steps clearly demonstrated individual differences in the typing rhythm and speed of users. In particular, significant differences were observed in the distribution of wavelet coefficients between typing samples belonging to different users [22].

The results of the analysis showed that low-level approximation coefficients reflected the general direction of the writing process, while high-level detail coefficients represented rapid changes over a short period of time. It was these detailing coefficients that became important in distinguishing users and made it possible to identify subtle and stable styles inherent in the writing process. As a result, it was found that features extracted based on the Xaar wavelet have high discrimination ability in user identification [19,20].

### 3.2 Test results of user identification algorithms

The feature set generated based on the Haar wavelet transform was tested using several machine learning algorithms. In particular, the K-Nearest Neighbors (KNN),

Support Vector Machine (SVM), and Random Forest (RF) models were used. During model evaluation, the data was divided into training and test sets, and each algorithm was tested under the same conditions [11,12,13].

The results of various classification algorithms were evaluated based on the criteria of Accuracy, Precision, Recall, and F1-score. The obtained results are presented in Table 1, which enables the comparison of model effectiveness and the identification of the most optimal algorithm.

**Table 1.** User identification results

| Algorithm | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| KNN | 86.7 | 0.88 | 0.87 | 0.87 |
| Random Forest | 90.1 | 0.90 | 0.90 | 0.90 |
| MLP | 91.5 | 0.92 | 0.92 | 0.92 |
| **SVM** | **92.6** | **0.93** | **0.93** | **0.93** |

As can be seen from the table, all algorithms showed high accuracy with features formed on the basis of the Haar wavelet transform. This case confirms that preserving the time-scale structure of keyboard typing dynamics significantly improves user recognition efficiency. In particular, a comparative analysis was performed in Table 2 based on the accuracy and balanced evaluation indicators F1-macro, F1-weighted, and F1-score of the KNN, Random Forest, MLP, and SVM models.

**Table 2.** Classification algorithm results (based on F1-score)

| Algorithm | Accuracy | F1-macro | F1-weighted | F1-score |
|---|---|---|---|---|
| KNN | 86.7 | 0.8680 | 0.8681 | 0.87 |
| Random Forest | 90.1 | 0.9015 | 0.9015 | 0.90 |
| MLP | 91.5 | 0.9158 | 0.9158 | 0.92 |
| **SVM** | **92.6** | **0.9272** | **0.9272** | **0.93** |

According to the results, the SVM algorithm achieved the highest accuracy rate (92.6%). This is explained by the fact that the SVM model can effectively form optimal splitting limits in the space of high-dimensional wavelet features [13]. The MLP (Neural Network) model also demonstrated high accuracy and proved effective in studying nonlinear connections [12,23]. While the Random Forest algorithm provided stable and balanced results, the KNN algorithm had a relatively lower accuracy due to the complexity of the feature space [11].

## 4 Discussion

The obtained experimental results showed that the effectiveness of classical machine learning algorithms in biometric identification systems is different. Among the four models studied during the experiment - SVM, MLP, Random Forest, and KNN - the

SVM (Support Vector Machine) algorithm showed the highest indicators in all metrics.

**Table 3.** Overall performance indicators of models

| Algorithm | Accuracy | F1-macro | F1-weighted | FAR | FRR | EER |
|---|---|---|---|---|---|---|
| KNN | 0.865236 | 0.868122 | 0.868122 | 0.013768 | 0.053137 | 0.033453 |
| Random Forest | 0.901167 | 0.900075 | 0.900075 | 0.023462 | 0.023529 | 0.023496 |
| MLP | 0.916085 | 0.915871 | 0.915872 | 0.015226 | 0.015686 | 0.015457 |
| **SVM** | **0.927647** | **0.927261** | **0.927262** | **0.012541** | **0.012353** | **0.012447** |

According to the data in Table 3, the SVM model significantly surpasses other models with an accuracy of 92.76%. The closeness of the F1-macro and F1-weighted (0.927) indicators confirms the model's robustness to class imbalance in the dataset and its high generalization ability. This once again proves the effectiveness of SVM in the isolation of biometric features in multidimensional space.

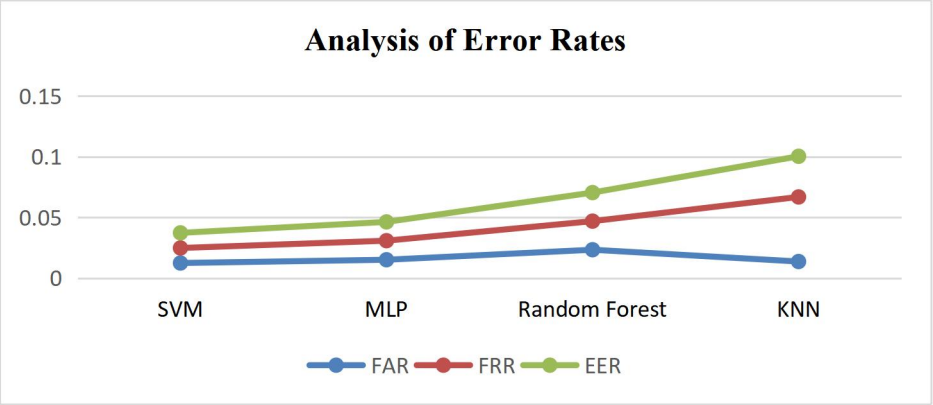



**Fig. 1.** Comparative diagram of the selected models by error levels (FAR, FRR, EER)

The effectiveness of classification models is determined not only by their overall accuracy, but also by the features of the errors made. Especially in biometric identification and security systems, the balance between FAR (False Acceptance Rate) and FRR (False Rejection Rate) is important (Fig. 1).

According to the results, the EER (Equal Error Rate) indicator in the SVM model was the lowest and amounted to 0.0124. The low EER value indicates that the model was able to find the optimal balance point between FAR and FRR. This effectively prevents unauthorized access to the system while avoiding unnecessary obstacles for legitimate users [24]. The MLP model (EER=0.0154) also showed high stability, confirming the ability of neural networks to reduce errors by learning nonlinear relationships.

Analysis of errors showed that the KNN model is the weakest link in this group. Its FRR (0.0531) is significantly higher compared to other models. A high FRR indicator causes inconvenience to system users, meaning the system often rejects genuine users.

This can be explained by the fact that the KNN algorithm is sensitive to data noise and makes more errors in boundary cases when making decisions based on the "nearest neighbor" principle.

Although the Random Forest model performed moderately well, its error rate was not as low as that of SVM. This means that the tree ensemble cannot accurately adapt to specific changes (variations) in the dataset, such as SVM [25].

## 5 Conclusion

In this study, a feature extraction approach based on Haar wavelet transformation was proposed and experimentally evaluated for user identification based on keyboard record dynamics. In the research process, the time sequences recorded by users during keyboard typing were subjected to Haar multilevel wavelet decomposition, and the obtained approximation and detail coefficients were directly used as symbols in the classification process.

Experimental results showed that features extracted based on the Xaar wavelet effectively reflect the time-scale structure inherent in users' writing process and significantly improve identification accuracy. As a result of tests using various machine learning algorithms, it was found that the SVM (RBF) and MLP (neural network) models achieved the highest accuracy and F1 scores, which is explained by the ability to deeply study a complex feature space consisting of wavelet coefficients.

Within the framework of further research, it is planned to introduce features based on Haar wavelet into deep learning models, increase the number of users, and assess the stability of the system under various input conditions (free text, different types of keyboards). The results of this work serve as a reliable scientific basis for the development of biometric authentication systems based on the dynamics of keyboard typing.

## References

1. Smith J. (2019). Modern Authentication Mechanisms: Security and Limitations. Journal of Cybersecurity Research, 12(3), 145–160.
2. Kumar, R., & Lee, J. (2020). Biometric Systems and Privacy Concerns. International Journal of Information Security, 8(1), 22–34.
3. Kevin S. Killourhy and Roy A. Maxion. "Comparing Anomaly Detectors for Keystroke Dynamics," in Proceedings of the 39th Annual International Conference on Dependable Systems and Networks (DSN-2009), pages 125-134, Estoril, Lisbon, Portugal, June 29-July 2, 2009. IEEE Computer Society Press, Los Alamitos, California, 2009.
4. T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer. "ROCR: visualizing classifier performance in R," *Bioinformatics* 21(20):3940-3941 (2005).
5. Mallat Stéphane, A Wavelet Tour of Signal Processing (Third Edition), Academic Press, 2009, Pages 377-434, ISBN 9780123743701, https://doi.org/10.1016/B978-0-12-374370-1.00012-4.
6. Kevin Killourhy and Roy Maxion. Keystroke dynamics-benchmark data set. Carnegie-MellonUniversity, http://www. cs. cmu. edu/˜ keystroke/# sec2, 2009.

7. S. Cho, C. Han, D. H. Han, and H. Kim. Web-based keystroke dynamics identity verification using neural network. Journal of Organizational Computing and Electronic Commerce, 10(4):295–307, 2000.
8. Yu Zhong, Yunbin Deng, and Anil K Jain. Keystroke dynamics for user authentication. In 2012 IEEE computer society conference on computer vision and pattern recognition workshops, pages 117–123. IEEE, 2012.
9. Yan Sun, Hayreddin Ceker, and Shambhu Upadhyaya. Shared keystroke dataset for continuous authentication. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2016.
10. Pranit Shinde, Saideep Shetty, and Mahendra Mehra. Survey of keystroke dynamics as a biometric for static authentication. International Journal of Computer Science and Information Security, 14(4):203, 2016.
11. Jiankun Hu, Don Gingrich, and Andy Sentosa. A k-nearest neighbor approach for user authentication through biometric keystroke dynamics. In 2008 IEEE International Conference on Communications,pages 1556–1560. IEEE, 2008.
12. Norhaslinda Harun, Satnam Singh Dlay, and Wai Lok Woo. Performance of keystroke biometrics authentication system using multilayer perceptron neural network (mlp nn). In *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*, pages 711–714. IEEE, 2010.
13. Romain Giot, Mohamad El-Abed, and Christophe Rosenberger. Keystroke dynamics with low constraints svm based passphrase enrollment. In 2009 IEEE 3rd International Conference on Biometrics:Theory, Applications, and Systems, pages 1–6. IEEE, 2009.
14. Alaa Darabseh and Akbar Siami Namin. Effective user authentications using keystroke dynamics based on feature selections. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pages 307–312. IEEE, 2015.
15. Ned Bakelman, John V Monaco, Sung-Hyuk Cha, and Charles C Tappert. Keystroke biometric studies on password and numeric keypad input. In 2013 European Intelligence and Security Informatics Conference, pages 204–207. IEEE, 2013.
16. Alsultan, A., & Warwick, K. (2013). Keystroke dynamics authentication: a survey of free-text methods. International Journal of Computer Science Issues (IJCSI), 10(4), 1.
17. Radomir S. Stanković, Bogdan J. Falkowski, The Haar wavelet transform: its status and achievements, Computers & Electrical Engineering, Volume 29, Issue 1, 2003,Pages 25-44, ISSN 0045-7906, https://doi.org/10.1016/S0045-7906(01)00011-8.
18. D. Sundararajan, "The Haar Discrete Wavelet Transform," in *Discrete Wavelet Transform: A Signal Processing Approach* , Wiley, 2015, pp.97-130, doi: 10.1002/9781119113119.ch8.
19. Mirzaev. N., Urinboev. J., Gafforov. N., Nugmanova. M. Analysis of Speech Signal Processing Methods in Speech Recognition Systems. 2025 6th International Conference on Problems of Cybernetics and Informatics (PCI), Baku, Azerbaijan, 26-28 August 2025, pp.
20. Opanasenko, V. M., Fazilov, S. K., Mirzaev, N. M., & Kakharov, S. S. (2025). A Model of Recognition Algorithms Based on Threshold Functions and Object Proximity Assessment. Cybernetics and Systems Analysis, 61(1), 148-160.
21. E. Yu and S. Cho. GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), pages 2253–2257. IEEE Press, 2003.
22. B. Hwang and S. Cho. Characteristics of auto-associative MLP as a novelty detector. In Proceedings of the IEEE International Joint Conference on Neural Networks, volume 5, pages 3086–3091, 10–16 July, 1999, Washington, DC, 1999.

23. S. Cho, C. Han, D. H. Han, and H. Kim. Web-based keystroke dynamics identity verification using neural network. Journal of Organizational Computing and Electronic Commerce, 10(4):295–307, 2000.
24. R. S. Gaines, W. Lisowski, S. J. Press, and N. Shapiro. Authentication by keystroke timing: Some preliminary results. Technical Report R-2526-NSF, RAND Corporation, May 1980.
25. Ramzi Saifan, Asma Salem, Dema Zaidan, Andraws Swidan, et al. A survey of behavioral authentication using keystroke dynamics: Touch screens and mobile devices. Journal of Social Sciences (COES&RJ-JSS), 5(1):29–41, 2016.

# Intelligent Information-Measuring System for Scanning Liquid Flows

Heydar Suleyman Rzayev[1][0009-0009-0335-2865], Arzu Agababa Akhundov[1][0009-0007-7946-8333], Natavan Mammad Khasayeva[2][0000−0003−3496−2040], Mehriban Bakhtiyar Mammadova[3][0000-0002-2362-222X]

[1]Institute of Mathematics, B.Vahabzadeh, 9. 1141, Baku, Azerbaijan

[2]Azerbaijan Technical University, Baku, Azerbaijan

[3]Azerbaijan State University of Oil and Industry, Baku, Azerbaijan

rzaaeeff@gmail.com; akhundovarzu@gmail.com; nxasayeva1@gmail.com; mehriban_m_1983@mail.ru

**Abstract.** This study presents the development and validation of an Automated Liquid Flow Scanning and Intelligent Measurement System (ALFSS) designed to provide high-accuracy, adaptive, and real-time monitoring of liquid flow processes in complex industrial and scientific environments. Traditional liquid flow measurement systems, including mechanical flow meters, analog sensors, and conventional digital measurement platforms, often suffer from limitations such as low adaptability, high sensitivity to noise, calibration drift, and reduced reliability in nonlinear and dynamically changing flow conditions. These limitations significantly affect the stability, accuracy, and robustness of measurement results, particularly in large-scale industrial infrastructures and safety-critical systems.

To address these challenges, the proposed system integrates multisensor measurement architectures**,** adaptive filtering techniques**,** intelligent data processing and real-time decision-making algorithms into a unified intelligent measurement platform. The system architecture is based on a layered structure consisting of a sensor layer, signal conditioning layer, data acquisition layer, processing unit, intelligent analysis layer, decision system, and visualization and control interface. This modular design ensures scalability, flexibility, and compatibility with different industrial and laboratory environments.

A comprehensive mathematical modeling framework is developed, including fluid flow dynamics models, state-space representations, measurement equations, and integrated error models. The system incorporates stochastic modeling approaches to represent uncertainties, sensor noise, calibration errors, and environmental disturbances. Adaptive estimation is achieved through Kalman filtering and extended data fusion techniques, enabling optimal state estimation under noisy and uncertain conditions. Multisensor data fusion strategies are implemented to enhance measurement reliability, improve accuracy, and provide redundancy against sensor failures.

An intelligent algorithmic framework is proposed, including automated flow scanning, adaptive filtering, sensor fusion, fault detection, anomaly analysis, and intelligent decision generation. The system is capable of detecting abnormal flow behaviors, identifying system faults, and generating corrective control actions in real time. Simulation models are developed using MATLAB/Simulink environments to validate system performance, stability, and robustness. Performance evaluation is carried out using quantitative metrics such as root mean square error (RMSE), accuracy indices, response time, and stability criteria based on Lyapunov theory.

The obtained results demonstrate that the proposed ALFSS significantly outperforms classical measurement systems in terms of accuracy, adaptability, robustness, and reliability. The system shows strong resistance to noise, high sensitivity to dynamic flow variations, and stable performance under nonlinear operating conditions. The developed framework is suitable for applications in water supply systems, oil and gas pipelines, chemical process industries, energy systems, hydraulic infrastructures, laboratory measurement platforms, and aerospace fluid systems.

The proposed smart measurement system provides a scalable, adaptive, and future-oriented solution for next-generation automated fluid flow monitoring and control technologies, providing higher accuracy, stability, and adaptability compared to classical measurement methods [1–5].

**Keywords:** Intelligent, measurement system, fluid flow, scanning, multisensor, test measurement, Kalman filter, adaptive control

# 1.Introduction

Real-time measurement and control of fluid flows in modern industrial and technological systems is an important scientific and technical problem. Classical measurement systems (mechanical flowmeters, analog sensors, differential pressure measurement methods) show low accuracy, high noise effect and poor adaptability in nonlinear dynamic environments [6–9].

For this reason, the concept of intelligent measurement systems (IMS) has been developed [1, 3, 10]. The Automated Liquid Flow Scanning (ALFS) approach forms a complex system approach that provides not only measurement, but also structural analysis of the flow, tracking of change trajectories, anomaly detection and prediction [2, 4, 11].

Accurate, reliable and real-time measurement of fluid flows in modern industry, energy, water management, chemistry, oil and gas and hydraulic systems remains one of the important scientific and technical problems. Nonlinear variability of flow parameters, turbulent regimes, influence of environmental factors, sensor errors and noise components seriously limit the capabilities of classical measurement systems [6–9]. Traditional mechanical flowmeters, analog sensor platforms and simple digital measuring devices do not provide stable results in complex dynamic environments, create high calibration errors and reveal long-term reliability problems [22].

**2.The physical basis of the flow measurement process**

The physical basis of the flow measurement process is based on the laws of classical hydrodynamics and is presented in a simple form by the following expression [6,7]:

$$Q = A \cdot v, \tag{1}$$

where $Q$ – is the flow rate, $A$ – is the cross-sectional area, and $v$ – is the flow velocity. However, in real industrial systems, this simple model is not sufficient, and the process is described by more complex dynamic structures based on the Navier–Stokes equations [17,18]:

$$\rho\left(\frac{\partial v}{\partial t} + v \cdot \nabla v\right) = -\nabla p + \mu\nabla^2 v + f, \tag{2}$$

This shows that the measurement process requires not only physical, but also mathematical modeling, computational methods, and intellectual analysis mechanisms.

Therefore, the multi-sensor intelligent information-measuring system we propose will provide high accuracy and reliability through the dual information it collects from sensors, special calculation methods and intelligent analysis mechanisms based on mathematical models [36].

*Intelligent Measurement Systems*

As a result of the transformation in measurement technologies in recent decades, the concept of Intelligent Measurement Systems (IMS) has been formed [1–5]. Unlike classical measurement structures, these systems not only integrate the measurement function, but also functions such as:

- adaptive analysis,
- self-calibration,
- noise immunity,
- decision-making,
- prediction,
- risk analysis,
- security mechanisms [10–15].

*Automated Liquid Flow Scanning*

In this context, the concept of Automated Liquid Flow Scanning (ALFS) is emerging as a new generation of measurement technology. The ALFS approach represents a complex intelligent system approach that integrates not only flow measurement, but also flow structure scanning, dynamic behavior modeling, anomaly detection, and adaptive control [11, 13].

In modern measurement systems, the measurement process is no longer limited to simple sensor readings and is expressed by the following mathematical structure [19–21]:

$$z(t) = h\big(x(t)\big) + \varepsilon(t), \tag{3}$$

where $z(t)$ - the measured signal, $x(t)$ - the real state, $h(\cdot)$ - the measurement function, $\varepsilon(t)$ - the measurement error. This model shows that the measurement process is uncertain and stochastic.

Therefore, state-space models are used in modern systems [20, 21]:

$$\dot{x}(t) = Ax(t) + Bu(t) + w(t), \tag{4}$$

$$y(t) = Cx(t) + v(t), \tag{5}$$

This approach allows for an integrated modeling of the dynamic behavior of the system, external influences, and measurement errors.

In addition, the Kalman filter and its extended forms form the main mathematical basis of adaptive measurement systems [1, 25, 26]. These filtering mechanisms provide optimal situation assessment in noisy conditions and significantly increase measurement accuracy.

In modern approaches, sensor fusion technologies are widely used [13, 27, 28]. This approach is expressed by the following model:

$$\hat{x}_{fusion} = \sum_{i=1}^{n} w_i x_i, \ \sum_{i=1}^{n} w_i = 1, \tag{6}$$

This model increases the reliability of measurements, provides resistance to sensor failures, and increases the stability of the measurement system.

Thus, modern measurement systems are no longer simple measurement platforms, but are formed as intelligent information systems. In these systems, measurement, analysis, decision-making, and control are combined into a single integrated platform [12–15].

The purpose of this research is to develop a scientifically based, adaptive, multi-sensor and intelligent measurement architecture for Automated Liquid Flow Scanning / Intelligent Measurement System (ALFSS), mathematical modeling, construction of algorithmic structures, verification in a simulation environment and scientific justification of application possibilities. The proposed system, unlike classical measurement approaches, aims to ensure not only measurement accuracy, but also system adaptability, stability, security and intelligent decision-making capabilities.

This approach elevates ALFSS to the level of an intelligent technological platform, not just a technical measurement device, and forms it as a scientific model of the next generation of automated measurement and control systems.

**3. System Architecture**

The system is built on a multi-layered intellectual structure [12–14]:

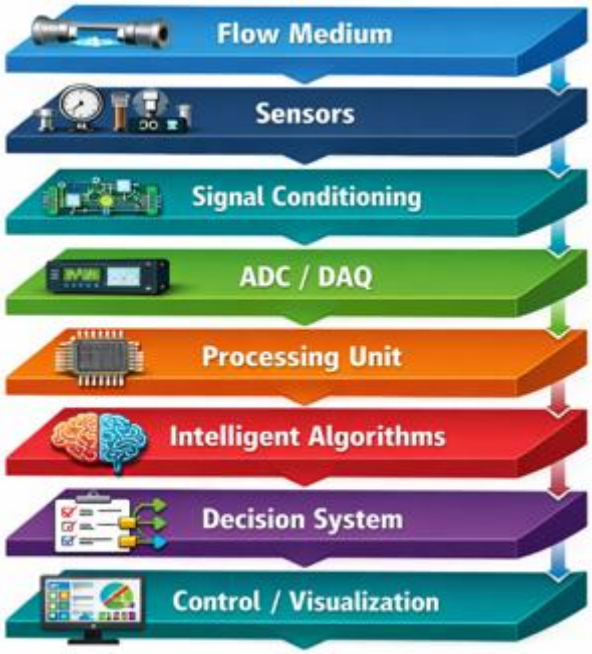



**Fig. 1.** Multi-level system architecture

This structure is based on the integration of sensor networks, real-time processing, and decision-making systems [13,15].

Thus, the total error as a result of complex measurements will be equal to the sum of all errors that must be taken into account and will be calculated by the following formula:

$$E_{total} = E_{sensor} + E_{noise} + E_{calibration} + E_{quantizion}, \quad (7)$$

This model reflects the integration of error components in complex measurement systems [22].

$E_{sensor}$ – Sensor Error: This error arises from the non-ideality of the physical sensor that directly performs the measurement. Causes:

- Sensor sensitivity is not ideal;
- Temperature effect;
- Nonlinearity;
- Hysteresis effect;
- Drift (change in parameters over time).

Mathematically, the real output of a sensor with a linear characteristic is expressed as:

$$y = kx + b + \varepsilon_s \quad , \quad (8)$$

where: $k$ – sensitivity error,

$b$ – offset,

$\varepsilon_s$ – sensor error.

Practical example: If the pressure sensor shows 102 kPa instead of 100 kPa → this is $E_{sensor}$ sensor.

$E_{noise}$ – Noise Error: This error is the random electrical or environmental effects added to the measurement signal. Sources:

- electromagnetic interference (EMI);
- thermal noise;
- internal fluctuations of electronic components;
- parasitic effects on transmission lines.

Characteristic:

- is random;
- its average value is usually close to zero:

$$E_{noise} \sim N(0, \sigma^2).$$

Noise reduction methods include:

- filtering (Kalman, low-pass filter);
- shielding;
- averaging.

$E_{calibration}$ – Calibration Error: Arise as a result of the sensor or measurement system not being properly adjusted to the reference values. Causes:

- incorrect calibration factor;
- outdated calibration;
- reference device error;
- environmental change.

Mathematical expression - If the real conversion is as follows:

$$y = k_{true}x,$$

but the system:

$$y = k_{cal}x,$$

then the system will function with the following expression :

$$E_{calibration} = \left(k_{cal} - k_{true}\right)x$$

Conclusion: This error is a systematic error and affects all measurements in the same direction.

$E_{quantizion}$ − Quantization Error: This error occurs when an analog signal is converted to discrete levels by an ADC (Analog-to-Digital Converter).

Cause: The ADC only accepts a limited number of levels: $2^N$ – where $N$ is the number of bits.

The quantization step is defined as follows:

$$\Delta = \frac{V_{max} - V_{min}}{2^N}$$

The maximum error will be determined by the following formula:

$$E_{quantizion} \leq \frac{\Delta}{2}$$

Example: 10-bit ADC → 1024 levels. The signal is rounded to the nearest digit → information loss occurs.

Physical meaning: The measurement process is the following chain process:

**Sensor → Electronics → ADC → Processing → Result**

Each stage of the functional modules adds its error and the result is $E_{total}$.

**Table 1.** Short comparison table

| Error type | Character | Source | Reduction method |
|---|---|---|---|
| $(E_{sensor})$ | Systematic | Sensor physical property | High quality sensor |
| $(E_{noise})$ | Random | Electrical noise | Filtration |
| $(E_{calibration})$ | Systematic | Miscalibration | Periodic calibration |
| $(E_{quantization}$ | Discretization | ADC bit limitation | High bit ADC |

## 4. Algorithms

4.1. Flow scanning algorithm

$$\begin{aligned} x_k &= f(x_{k-1}, u_k) + \omega_k \\ y_k &= h(x_k) + v_k \end{aligned} \tag{9}$$

This model forms the basis of discrete adaptive measurement systems [23, 24].

Kalman filtering

$$\begin{aligned} K_k &= P_k H^T (HP_k H^T + R)^{-1} \\ \hat{x}_k &= \hat{x}_{k-1} + K_k(y_k - H\hat{x}_{k-1}) \\ P_k &= (1 - K_k H)P_{k-1} \end{aligned} \tag{10}$$

The Kalman filter is considered an optimal filtering method in adaptive measurement systems [1, 25, 26].

Sensor fusion

$$\hat{x}_{fusion} = \sum_{i=1}^{n} \omega_i x_i$$

$$\sum_{i=1}^{n} \omega_i = 1$$

This approach is used to increase accuracy in multisensor systems [27, 28].

Fault detection$r_k = y_k - \hat{y}_k$

$$|r_k| > \delta \Rightarrow \text{Fault detected}$$

Anomaly detection is a key component of security-based systems [29, 30].

**4.2. Modeling and Statistical Estimation**

The root mean square error (RMSE) formula is a widely used measure of the residual standard deviations and also a measure of forecasting error. It estimates the average difference between the values predicted by a model and the actual observed values. It is a measure of confounding factors, meaning smaller values indicate a better fit between the model and the data. RMSE is very sensitive to outliers because it squares the deviations before averaging. Performance indicators [31-33]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k - \hat{x}_k)^2}$$

$$Accuracy = 1 - \frac{|x - \hat{x}_k|}{x}$$

Lyapunov Stability Criterion [21,34]

$$V(x) > 0,\ \dot{V}(x) < 0$$

**4.3. Test measurement algorithm for scanning parameter values**

A test measurement method is proposed to determine the measurement errors of sensors with linear or nonlinear conversion characteristics, thanks to which additive and multiplicative tests are of particular importance [35,36].

As an example, simultaneous measurement of flow rate, volume, and temperature in a pipeline is considered. With the help of the system we offer, it is possible to read the current readings of smart sensors installed on any pipeline. If we assume a quadratic transformation function:

$$f_x = b_0 + b_1\rho_x + b_2\rho_x^2, \quad (12)$$

where $f$ – the output quantity of the transmitter, frequency;

$\rho_x$ – the input quantity of the transmitter, measurement parameter, density;

$b_0, b_1, b_2$– the parameters, coefficients of the transmitter's TF.

This algorithm, consisting of four measurement cycles, allows for highly accurate determination of the coefficients of the conversion characteristic. The generalized result of measurements performed with combined multiplicative and additive tests is determined by the following formula:

$$\rho_x = \frac{(f_0 - f_3) + (f_1 - f_2)}{(f_0 - f_3) - (f_1 - f_2)}\rho_{et}, \tag{13}$$

Since the errors of the measurement cycles are taken into account, the difference of these expressions will give the error ($\Delta_T$) of the tested MS:

$$\Delta_T = (\rho_x + \rho_{et})(\Delta_1 - \Delta_2) + (\rho_1 - \rho_{et})(\Delta_3 - \Delta_0), \tag{14}$$

By substituting the calculated values of the additive and multiplicative tests, as well as the measurement errors, into the function that determines the final error of the test MS, we obtain the following mathematical model for the total (total) error:

$$\Delta_T = \rho_{et}[(\Delta_1 - \Delta_2) - (\Delta_3 - \Delta_0)] + \rho_x[(\Delta_1 - \Delta_2) + (\Delta_3 - \Delta_0)]. \tag{15}$$

We obtain the following expression for the absolute error $\Delta_{inp.}$ introduced into the input of the tested MS:

$$\Delta_{inp.} = f_T^{-1}[f_T(x) + \Delta_T] - x = \frac{\Delta_T}{f(x)_T^{-1}}, \tag{16}$$

where, $f_T(x) = (y_3 - y_0)(\rho_x - \rho_{et}) + (y_2 - y_1)(\rho_x + \rho_{et})$,

If we differentiate expression with respect to $x$, we get the following value for

$$f'_T(x) = (y_0 - y_3) - (y_1 - y_2), \tag{17}$$

If we substitute expressions (15) and (18) into (17), we get the following expression:

$$\Delta_{inp.} = \frac{\rho_{et}[(\Delta_1 - \Delta_2) - (\Delta_3 - \Delta_0)] + \rho_x[(\Delta_1 - \Delta_2) + (\Delta_3 - \Delta_0)].}{2\rho_{et}\big(b_2 + b_3(\rho_x + \rho_{et})\big)}, \tag{18}$$

If the real transfer function of the initial MS is in the form of a third-degree polynomial, then its identification as a result of piecewise nonlinear approximation is also realized by the appropriate algorithm.

Thus, all error components are random variables, and the distribution laws of $\Delta_i$ are defined as the sum of the distributions of random variables and, therefore, can be considered normal. Therefore, for their complete description, it is enough to know the MΔ -mathematical expectation and DΔ -variance.

From this we can conclude that under the mentioned conditions, the main characteristics of the errors of the tested measurement, which implement the algorithms for increasing the measurement accuracy of the MS based on simple

additive and multiplicative tests, are justified for the tested IMS operating on the basis of the optimal set of additive, multiplicative and combination tests.

When the errors of the measurement cycles are independent of each other, due to the application of the combined test algorithm, we obtain the following expression for the variances of the errors of the tested MS:

$$\sigma^2_{\Delta_T} = \sigma^2_{\Delta_0}[x(k-1)-\theta)]2 + \sigma^2_{\Delta_1}[x(k-1)+\theta] + \sigma^2_{\Delta_2}[x(k-1)+\theta] + \sigma^2_{\Delta_0}[x(k-1)-\theta)], \quad (19)$$

As can be seen from the expressions obtained above, the dispersion of the $\Delta_T$ error of the main test equations of the MS tested in the algorithmic-test method of increasing measurement accuracy is increased compared to the dispersion of a single-step measurement.

As can be seen, the uncorrelated components of the additive error generally prevail, so the coefficient $K^*_{\sigma_{ad.}}$ depends on the values of the additive and multiplicative tests.

Clarify, the dependence of the coefficient $K^*_{\sigma_{ad.}}$ on the values of additive and multiplicative tests: as can be seen from expression, the detected correlation relation (CR) between the measured value, additive, and multiplicative test parameters is visually reflected by the presence of their values in the area $A_0$.

As a result of the application of algorithm, the uncorrelated components of the additive error of the TMS measurement result do not increase, at least because of the interrelation of the same error components. Graph 1 describes the dependence of $K^*_{\sigma_{ad.}}(\theta,k)$ in case of $x=x_0$.

As can be seen from the Graph, the dependence $K^*_{\sigma_{ad.}}(\theta,k)$, a given range of values of additive and multiplicative tests, is located behind the line intersecting with the points $A^*A^*$ on the the single-level plane $K^*_{\sigma_{ad.}}$.

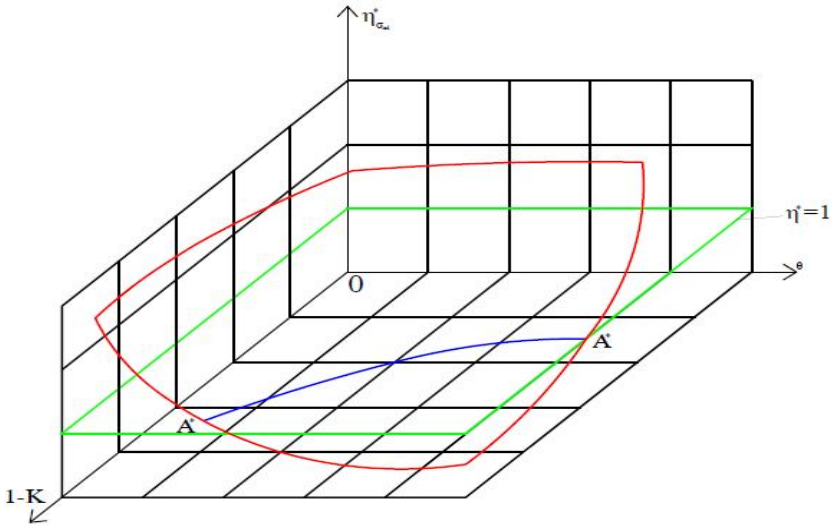
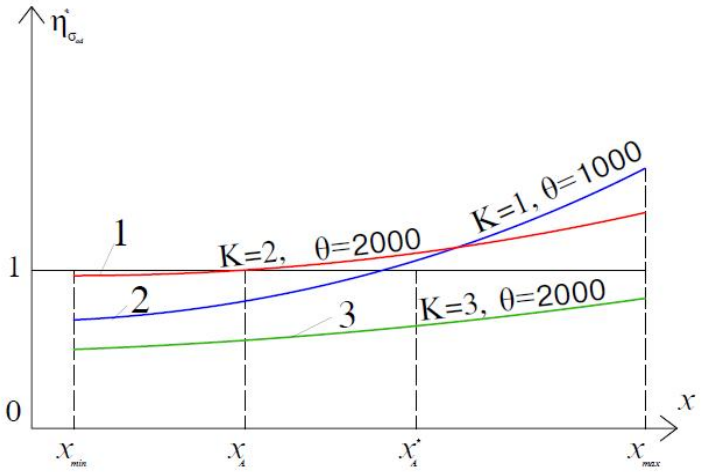



**Fig. 2.** Dependency graph $K^*_{\sigma_{ad.}}(\theta,k)$

Fig. 3 illustrates the dependence of $K^*_{\sigma_{ad.}}(x)$ on the given values of the additive and multiplicative tests selected from area $A_0$ . When the $x$ values of areas $\theta$ and $k$ are given, the additive components of the measurement error occurring during the execution of algorithm (90) do not increase compared to the analogous IMS error, are limited to values $x_A$, $x^*_A$ and .The curve 3 corresponding to the $\theta$ and $k$ values define the all IMS measurement range in this area.




Graph 2. Dependency graph $K^*_{\sigma_{ad.}}(x)$

here, 1 – dependency $K^*_{\sigma_{ad.}}(x)$, in case of $k=2,\ \ \theta=2000$ ; 2 – dependency $K^*_{\sigma_{ad.}}(x)$, in case of $k=3,\ \ \theta=1000$ ; 3 – dependency $K^*_{\sigma_{ad.}}(x)$, in case of $k=3,\ \ \theta=2000$

**Conclusion.** Thus, the study of the metrological characteristics of the tested MS shows that the following factors most affect the resulting measurement error of this type of systems:

– the factor of the error arising from the use of fixed factors of additive and multiplicative tests;

– the uncorrelated factor of the static error of the tested MS;

– the factor of the dynamic error of the tested MS;

– the factor of the error arising from the inadequacy of the adopted mathematical model of the real conversion characteristic of the tested MS.

The developed test algorithms were tested on a real automatic density measuring (ADM-2M) device sample, the measurement errors were evaluated, and a decrease in the total error by several times was confirmed.

## References


1. Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *ASME Journal of Basic Engineering*, 82(1), pp. 35–45.
2. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, Germany.
3. Haykin, S. (2009). *Neural Networks and Learning Machines*. Pearson, USA.
4. Simon, D. (2006). *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*. Wiley, USA.
5. Åström, K. J., Murray, R. M. (2010). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, USA.
6. White, F. M. (2011). *Fluid Mechanics*. McGraw-Hill, USA.
7. Fox, R. W., McDonald, A. T., Pritchard, P. J. (2015). *Introduction to Fluid Mechanics*. Wiley, USA.
8. Çengel, Y. A., Cimbala, J. M. (2014). *Fluid Mechanics: Fundamentals and Applications*. McGraw-Hill, USA.
9. Benedict, R. P. (1984). *Fundamentals of Pipe Flow*. Wiley, USA.
10. Frank, P. M. (1990). Fault diagnosis in dynamic systems. *Automatica*, Elsevier, 26(3), pp. 459–474.
11. Qin, S. J., Badgwell, T. A. (2003). A survey of industrial model predictive control. *Control Engineering Practice*, Elsevier, 11(7), pp. 733–764.
12. Isermann, R. (2005). Model-based fault-detection and diagnosis. *Springer*, Germany.
13. Luo, R. C., Kay, M. G. (1989). Multisensor integration and fusion. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5), pp. 901–931.
14. Hall, D. L., Llinas, J. (2001). *Handbook of Multisensor Data Fusion*. CRC Press, USA.
15. Varshney, P. K. (1997). *Distributed Detection and Data Fusion*. Springer, USA.
16. Streeter, V. L., Wylie, E. B., Bedford, K. W. (1998). *Fluid Mechanics*. McGraw-Hill, USA.
17. Batchelor, G. K. (2000). *An Introduction to Fluid Dynamics*. Cambridge University Press, UK.
18. Landau, L. D., Lifshitz, E. M. (1987). *Fluid Mechanics*. Pergamon Press.
19. Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, USA.
20. Ogata, K. (2010). *Modern Control Engineering*. Prentice Hall, USA.
21. Khalil, H. K. (2002). *Nonlinear Systems*. Prentice Hall, USA.
22. Doebelin, E. O. (2004). *Measurement Systems: Application and Design*. McGraw-Hill, USA.
23. Maybeck, P. S. (1979). *Stochastic Models, Estimation and Control*. Academic Press, USA.
24. Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, USA.

25. Welch, G., Bishop, G. (2006). An Introduction to the Kalman Filter. UNC Chapel Hill Report.
26. Grewal, M. S., Andrews, A. P. (2015). *Kalman Filtering: Theory and Practice*. Wiley, USA.
27. Khaleghi, B., Khamis, A., Karray, F. O., Razavi, S. N. (2013). Multisensor data fusion. *Information Fusion*, Elsevier, 14(1), pp. 28–44.
28. Elmenreich, W. (2002). Sensor fusion in time-triggered systems. *PhD Thesis*, Vienna University of Technology.
29. Venkatasubramanian, V. et al. (2003). A review of process fault detection. *Computers & Chemical Engineering*, Elsevier.
30. Ding, S. X. (2008). *Model-Based Fault Diagnosis Techniques*. Springer, Germany.
31. Montgomery, D. C. (2012). *Introduction to Statistical Quality Control*. Wiley, USA.
32. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2008). *Time Series Analysis*. Wiley, USA.
33. Nise, N. S. (2011). *Control Systems Engineering*. Wiley, USA.
34. Slotine, J. J. E., Li, W. (1991). *Applied Nonlinear Control*. Prentice Hall, USA.
35. Isaev M.M. Methods and means of commercial accounting of petroleum products in the flow. Baku: Elm Publishing House, 2010, 164 p.
36. Isayev M.M. Algorithmic-testing methods for increasing the accuracy of measurement systems. Baku, "Elm" publishing house, 2018, 206p.

# ICT and Intelligent Systems in Sociology and Economics

section

# Noise in Big Cities and Methods of Combating It

Doctor of Philosophy in Architecture, Associate Professor

S. I. Allakhverdieva[1][0000-0002-7501-5691]

Doctor of Philosophy in Architecture, Associate Professor

V. A. Jafarova[2][0009-0007-7023-0812]

Azerbaijan University of Architecture and Construction,
Baku, Azerbaijan

**Abstract.**

One of the most widespread harmful factors of the urban environment is noise. Noise sources interfere with people's rest, causing dissatisfaction and irritation. When designing buildings and structures, architects must take into account the noise load on individual facilities as well as on the entire building complex. Noise protection can be implemented both at the source of noise generation and along its path of propagation. For the effective adoption of appropriate measures, it is necessary to know the noise characteristics of the sources. Noise maps make it possible to more rationally plan measures to combat noise of both urban-planning and administrative nature related to the organization of urban transport traffic. Transport, industrial, and domestic noise are considered as the main noise sources. The effectiveness of protecting buildings from sound waves increases with distance; moreover, the efficiency of protection is enhanced by greening the space located between the noise source and the receiver.

Satisfactory protection results can be achieved when the green zone extends over more than 100 m. The construction of buildings with enclosed courtyard layouts in developments strictly parallel to the street should be considered an effective measure; however, such an arrangement affects the architectural and planning solution of the building. The main methods used to limit noise propagation include appropriate external and internal spatial planning, the installation of adequate sound insulation for the enclosing structures of residential and public buildings, and related measures.

**Keywords:** Noise sources; noise load; transport noise; industrial noise; domestic noise; protection efficiency; sound insulation of enclosing structures.

# Noise in big cities and methods of combating it.

**Introduction:** Noise is one of the most widespread harmful factors of the urban environment. In noise-control practice, this term refers to any sounds that disturb people. When designing buildings and structures, architects must take into account the noise load on individual facilities as well as on the entire building complex. Noise sources interfere with people's rest, causing dissatisfaction and irritation. The impact of noise on humans varies depending on its level, spectral composition, and duration of exposure:

- low-level noise, typical of auditoriums and assembly halls, generally hinders the perception of speech and music;
- medium-level noise, occurring for example in office and administrative spaces, leads to premature fatigue among employees;
- high-level noise in industrial premises causes rapid worker fatigue and negatively affects labor productivity.

For this reason, already at the initial stage of developing the planning structure of buildings, it is necessary to consider the sound insulation of building elements and to ensure that indoor sound levels do not exceed permissible sanitary standards by applying a комплекс set of measures to protect buildings from noise penetrating into interior spaces.

**1.1 Methodology:** Noise protection can be implemented either at the source of noise generation or along its path of propagation. To successfully select and apply appropriate measures, it is necessary to determine the noise characteristics of the sources.

Measures to reduce noise in urban development and buildings must be envisaged at all stages of design:

1. at the stage of regional planning, addressing the placement of airports and transport infrastructure;
2. at the stage of developing the city master plan, determining a planning structure that reduces the number of noise sources and limits the extent of their impact;
3. at the stage of designing residential districts, defining the spatial, volumetric, and planning composition of development and landscaping that effectively utilizes the noise-protective qualities of the urban environment.

In the process of forming the master plan, noise maps should be prepared for the arterial road network based on projected traffic flows.

Noise maps consist of contour lines plotted on the plan, illustrating the reduction of noise levels with increasing distance from major roads. Such maps enable a more rational planning of noise mitigation measures of both urban-planning and administrative nature related to the organization of urban transport traffic.

- In the design of cities and other settlements, it is necessary to ensure a clear functional zoning of the territory into residential, industrial, utility–warehouse, and external transport zones.

- In the design and assessment of external noise, consideration is given to building orientation, articulation, and form; the presence of windows and balconies; the construction of embankments and barriers; the organization of green and protective zones; access roads and traffic layouts; as well as the planning solutions of individual buildings.

- When locating industrial and utility–warehouse zones, compliance with permissible noise standards must be ensured.

Transport, industrial, and domestic noise are considered the main types of noise sources. Cities are characterized by numerous noise sources, which can be условно classified into two major groups: individual sources and composite sources consisting of multiple individual emitters. Individual noise sources include single vehicles, electrical transformers, intake or exhaust openings of ventilation systems, installations of industrial or power-generation facilities, and similar elements. Composite noise sources include traffic flows on streets and roads, train movements on railways, industrial enterprises with multiple noise-generating units, sports or playground facilities, and others. When developing master plans for new or реконструируемых cities, industrial zones should be located on the leeward side relative to residential areas, based on data on prevailing wind directions in different seasons of the year. The effectiveness of protecting buildings from sound waves increases with distance; moreover, the protective effect is enhanced when the space between the noise source and the receiver is landscaped. Satisfactory protection results can be achieved when the green zone extends over more than 100 m

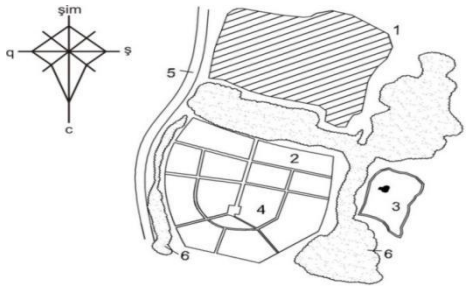



**Fig. 1. Diagram of the city master plan**

1 – industrial zone; 2 – green belt; 3 – recreational area; 4 – administrative and commercial center; 5 – main arterial road; 6 – protective zone. A number of methods have been developed for assessing various types of noise, many of which are widely applied in practice. In order to unify measurement and assessment methods for urban environmental noise, the international standard **ISO 1996-1 "Acoustics — Description, measurement and assessment of environmental noise.**

**Part 1: Basic quantities and assessment procedures"** was developed. According to this standard, the equivalent continuous sound level, expressed in **dBA**, should be used as the basic parameter for describing environmental noise conditions. An essential prerequisite for the correct evaluation of noise characteristics is the appropriate selection of the time interval over which the equivalent sound levels of noise sources are determined. The main cause of transport noise is street traffic. When designing street and road networks, terrain features should be utilized as natural barriers to noise propagation. Where it is necessary to route arterial streets and roads on embankments or elevated structures, the installation of noise barriers should be provided on the basis of acoustic calculations (Fig. 2). Artificial and natural terrain features (cuttings, embankments, hills, etc.) may also function as noise barriers. The reduction in sound level behind barriers is determined by the dimensions of the barrier, the distance between the barrier and the noise source, and other related factors. The effectiveness of screening walls or embankments should be verified through preliminary schematic design. In this approach, barriers are positioned close to the noise source in order to protect the open space located between the noise source and the receiver.

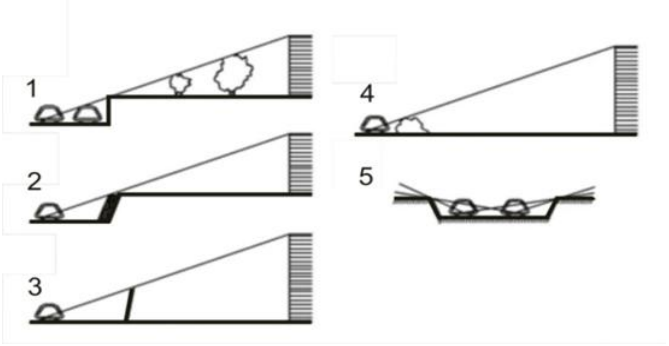



**Fig. 2. Schemes of noise barrier placement for the reduction of transport noise:**

1- wall;
2- inclined surface covered with sound-absorbing material;
3- noise-reducing screen;
4- mound/hill;
5- depressed roadway.

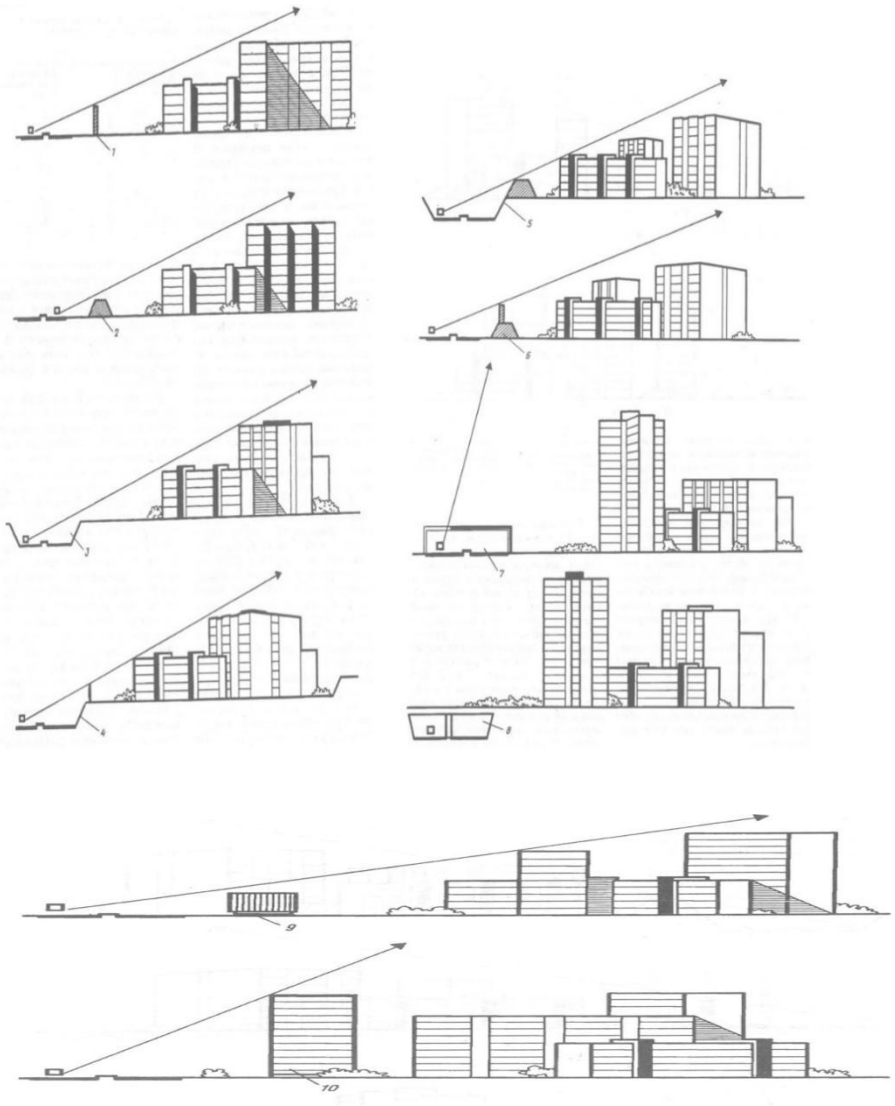


**Fig. 3. Types of noise barriers:**
1 – wall-type barrier; 2 – embankment barrier 3 – cutting (trench) barrier;
4 – terrace barrier; 5, 6 – combined barriers:
5 – cutting combined with an embankment or wall; 6 – embankment combined with a wall; 7 – gallery-type barrier; 8 – tunnel-type barrier; 9 – non-residential building used as a barrier 10 – noise-protective residential building.

Let us consider in more detail the means by which the above-mentioned noise reduction methods can be implemented. It is advisable to organize building layouts in such a way that all rooms associated with the generation of noise are concentrated in one area and located at a distance from working and residential spaces. Thus, in residential and public buildings, boiler rooms, elevator machine rooms, elevator shafts, refuse chutes, pumping stations, rooms housing fans, canteens, buffets, and similar facilities should not adjoin living or working areas. Kitchens, bathrooms, and sanitary facilities are recommended to be grouped into separate blocks adjoining stairwell walls or similar blocks of adjacent apartments. Living rooms in multi-storey residential buildings, dormitories, and hotels; workspaces in administrative buildings; wards in hospitals and sanatoriums; as well as classrooms and lecture halls in educational institutions should be separated from stairwells by auxiliary spaces (kitchens, bathrooms, corridors, etc.). Gymnasiums, workshops, and other noisy premises in educational buildings should not be located in close proximity to classrooms, lecture halls, or laboratories.

The construction of buildings with enclosed courtyard layouts in developments strictly parallel to the street should be regarded as an effective measure. However, such an arrangement influences the architectural and planning solution of the building. If buildings are constructed perpendicular to the street without additional screening measures, sound propagates freely, resulting in high noise levels around the building. When a residential building is used as a noise barrier, special measures are required to improve the acoustic environment within the building itself; such buildings are sometimes referred to as noise-protected buildings. Buildings may be designed with layouts in which the windows of all residential rooms face the courtyard façade, oriented toward the acoustic shadow behind the building. From the standpoint of noise protection, it is also advisable to locate commercial and service enterprises by attaching them to such building.

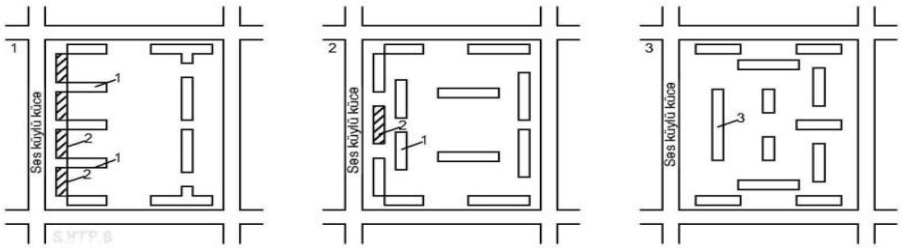



**Fig. 4. Optimal placement of residential buildings in relation to a noisy arterial road:**
1–residential buildings; 2-shopping centers and service buildings; 3 – sound-insulating barrier buildings.

One of the most effective ways to protect residential environments from transport noise is the construction of screening buildings (such as garages, business facilities, and commercial zones), as well as the development of barrier-type residential buildings with high sound-insulating performance along major transport corridors. Such buildings are sometimes referred to as **noise-protected buildings**.

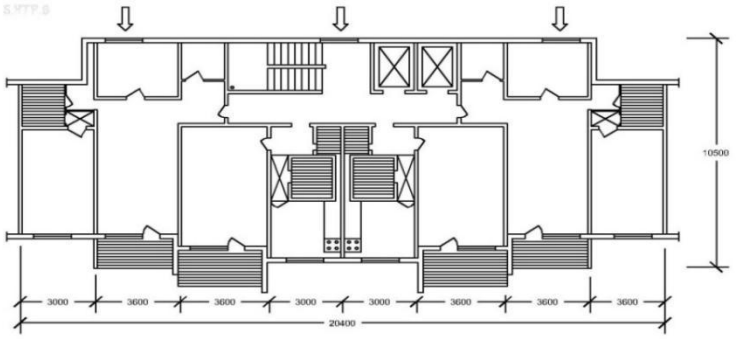



**Fig. 5. Four-apartment section on the side of non-residential premises of apartments in a 9–12-storey barrier-type residential building** ***(plan of a residential building with high sound-insulating performance).***

In urban development, green spaces may also be used for noise mitigation (Fig. 5).

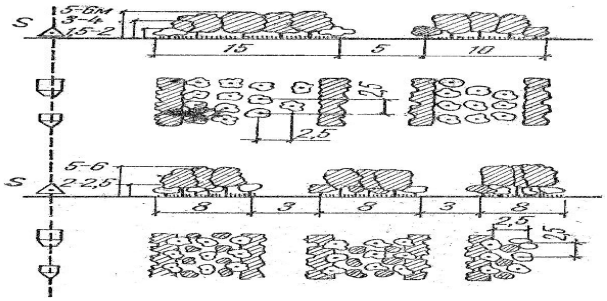



**Fig.6 Sound insulation with the help of green spaces**

Green spaces can be used to reduce noise in urban development**.**

In the noise environment of residential developments, so-called **intra-block noisesources** play a significant role. These include:

- sports and children's playgrounds within the courtyards of residential microdistricts;
- loading and unloading areas in service yards of shops, canteens, and other public catering and community service facilities;

- refuse collection vehicles, splash pools/fountains, garages, transformer substations, and similar facilities.

When organizing children's playgrounds within courtyards, in order to reduce sound reflection it is recommended to place them in the central part of the courtyard, where noise levels may be reduced by up to 3 dB.

**1.2 Noise sources in residential, public, and industrial buildings.**

Noise occurring in residential and public buildings can be classified into:

- **domestic noise**, associated with human activity;
- **mechanical noise**, related to the operation of engineering and sanitary equipment (elevators, fans, pumps, etc.).

Domestic noise is generated by residents or occupants of a building. Loud conversations, singing, playing musical instruments, children's shouting and crying, and especially the operation of televisions, radios, record players, and tape recorders are sources of airborne noise.

During walking, dancing, and the movement of furniture, sound vibrations arise in the building envelope; these vibrations are transmitted to floor slabs, walls, and partitions and propagate over long distances throughout the building in the form of structure-borne noise. This occurs due to the very low attenuation of sound energy in the materials commonly used for building structures.

Elevators installed in multi-story buildings are sources of significant noise, which propagates not only through the air within the shaft and stairwell, but primarily through the building structures as a result of the rigid attachment of the elevator shaft to walls and floor slab (Fig.7).

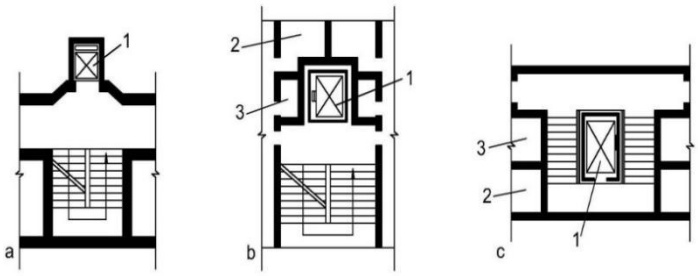



**Fig. 7. Optimal options for the placement of an elevator shaft within a stairwell:**

- the elevator shaft is located outside the stairwell;
- the elevator shaft is located within the building;
- the elevator shaft is positioned between the stair flights and the landing;

1 – elevator shaft;
2 – kitchen;
3 – sanitary unit.

Fans, pumps, elevator winches, and other mechanical equipment in buildings are sources of both airborne and structure-borne noise. Ventilation systems, in particular, generate intense airborne noise which, if appropriate measures are not taken, propagates along with the air flow through ventilation ducts and penetrates rooms through air grilles. Depending on the type of noise, different mitigation measures are adopted to reduce its propagation. The main methods used to limit noise transmission include appropriate external and internal spatial planning, the provision of adequate sound insulation for the enclosing structures of residential and public buildings, and related measures. propagation. The main methods used to limit noise transmission include appropriate external and internal spatial planning, the provision of adequate sound insulation for the enclosing structures of residential and public buildings, and related measures. propagation. The main methods used to limit noise transmission include appropriate external and internal spatial planning, the provision of adequate sound insulation for the enclosing structures of residential and public buildings, and related measures. Especially high noise levels may occur in rooms located directly beneath installed propagation. The main methods used to limit noise transmission include appropriate external and internal spatial planning, the provision of adequate sound insulation for the enclosing structures of residential and public buildings, and related measures. Fans, due to vibration, generate highly intense sound vibrations in floors and walls, which propagate easily as structure-borne noise throughout the building and are radiated into rooms, even those located far from the noise source (Fig. 8).

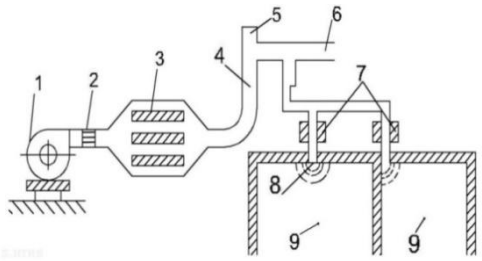



**Fig. 8. Noise control in ventilation systems:**

1 – fan or motor mounted on a vibration-isolation layer; 2 – rubber-covered canvas gasket; 3 – primary sound absorber; 4 – air duct; 5 – air mass; 6 – duct branch; 7 – additional sound absorber; 8 – ceiling panel; 9 – ventilated room.

Fans, due to vibration, generate highly intense sound vibrations in floors and walls, which propagate easily as structure-borne noise throughout the building and are radiated into rooms, even those located far from the noise source (Fig. 7). Ventilation systems and pumps are often located in basement rooms. In such cases, barriers are installed close to the noise source to protect the open space located between the source and the receiver.

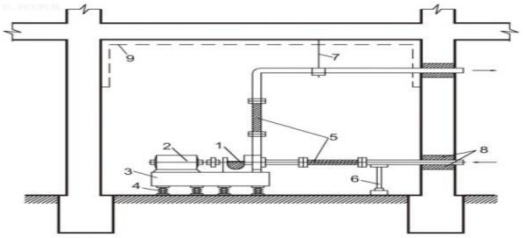



**Fig. 9. Sound insulation of a pump installation:**

1 – pump unit;
2 – motor;
3 – reinforced concrete base;
4 – vibration isolator;
5 – rubber covering;
6 – support;
7 – suspension;
8 – pipe vibration isolation;
9 – sound-absorbing material.

**Conclusions:** The primary means of protecting the interiors of residential and public buildings from noise is the proper sound insulation of enclosing structures, which must ensure compliance with regulatory requirements for soundproofing. In many rooms of public buildings, the installation of sound-absorbing wall finishes is advisable, for example in elongated spaces such as corridors in schools, hospitals, and hotels, as this prevents noise propagation along these spaces. To reduce noise in typing rooms, accounting offices, computing centers, administrative offices, restaurants, railway and airport waiting halls, shops, canteens, and similar facilities, sound-absorbing finishes for walls and ceilings should be provided. In most ventilation systems of public buildings, the use of **noise silencers** is necessary (Fig. 10).

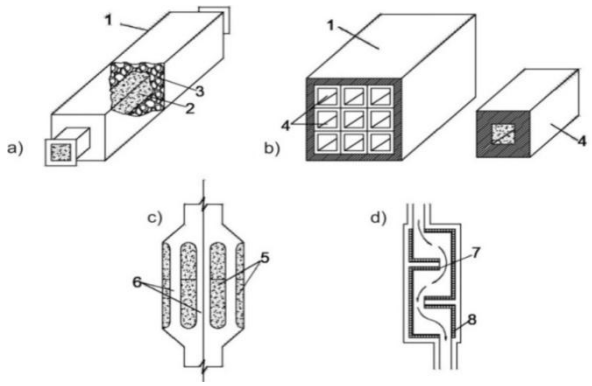


**Fig. 10. Schemes of noise silencers:**

a – tubular; b – cellular; c – plate or slit-type; d – chamber-type.

1 – outer shell; 2 – porous duct; 3 – sound-absorbing material;
4 – sound-absorbing cells; 5 – panels made of sound-absorbing material;
6 – duct; 7 – transverse partitions; 8 – sound-absorbing material.

**Industrial noise** differs from other types of noise in terms of its high level and spectral composition. The effectiveness of isolation depends on the noise spectrum, the technological process, and the equipment used. In production workshops with noisy equipment, isolation is required through special acoustic screens, partitions, ceiling structures, and similar constructions. These may include U- and L-shaped acoustic screens or freestanding panels.

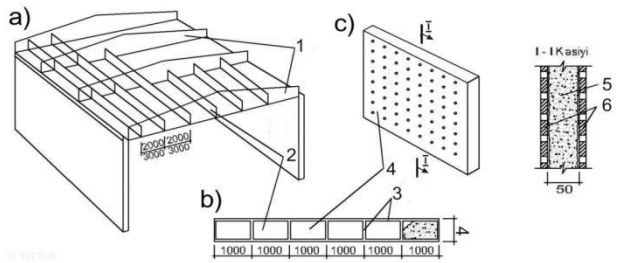



**Fig. 11. Arrangement and construction of a sound-absorbing beam:**

a – beam placement schemes; b – general view; c – sound-absorbing element.

1 – beam; 2 – sound-absorbing beam; 3 – beam frame; 4 – filler; 5 – mineral wool mats; 6 – porous metal element.

In the planning of industrial buildings, noisy rooms should be grouped into a single unit, reliably isolated from quiet areas by solid walls or auxiliary spaces (cloakrooms, sanitary facilities, airlock corridors, etc.). To reduce noise, technical solutions such as those shown in Fig. 11are applied. Technological processes in very noisy rooms are often carried out from **sound-insulated booths**

## References


1. Гусев Н.М. «Основы строительной физики». Учебное пособие.М.(1975).
2. Римша А.Н. «Градостроительство в условиях жаркого климата» М.(1979).
3. Коврыгин С.Д. «Архитектурно-строительная акустика» М.(1980).
4. Р.Г.Абдулрагимов «Salonların akusiği və tasarımı», Trabzon (2005).
5. Р.Г.Абдулрагимов «Mimarlıkta aydınlatma», Trabzon (2005).
6. Azərbaycan Respublikası Prezidentinin 8 iyul 2008-ci il tarixli Fərmanı "Ətraf mühitə və insan sağlamlığına mənfi təsir göstərən vibrasiya və səs-küy çirklənmələri normaları". Normativ sənəd, Bakı, (2008)
7. Бродич М. И др. «Инженерное оборудование высотных зданий» «АВОК-ПРЕСС». Учебное пособие.М.(2011)
8. .Г.Абдулрагимов,S.İ.Allahverdiyeva «Memarlıq fizikasının əsasları», Bakı (2012).
9. Аллахвердиева, Г. Гасанова. Архитектурно-строительная акустика. Учебное пособие. Баку, (2012)
10. R.H.Əbdülrəhimov. Memarlıq fizikası. Dərslik. Bakı-(2015).
11. Краснов Д.К. «Методы борьбы с шумом в городе». Статья.(2023).

# Quantum Transfer Learning with Pretrained CNN Features for Multi-Class Medical Image Classification

Furkan Atban[1,2][0000-0002-1712-5155], Muhammed Yusuf Küçükkara[1,2][0000-0003-0600-3651], and Cüneyt Bayılmış[3][0000-0003-1058-7100]

[1] Department of Computer Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Turkiye furkanatban@subu.edu.tr, muhammedkucukkara@subu.edu.tr

[2] Department of Computer Engineering, Institute of Natural Sciences, Sakarya University, Sakarya, Turkiye

[3] Department of Computer Engineering, Faculty of Computer and Information Sciences, Sakarya University, Sakarya, Turkiye cbayilmis@sakarya.edu.tr

**Abstract.** The automated classification of Optical Coherence Tomography (OCT) images is critical for the early diagnosis of retinal diseases such as choroidal neovascularization, diabetic macular edema and drusen. Classical deep networks achieve high accuracy but often require millions of parameters and significant computational resources. We present a hybrid Quantum Transfer Learning (QTL) framework combining a fine-tuned ResNet34 encoder with a variational quantum circuit (VQC) classifier for multi-class OCT classification. The classical backbone is partially fine-tuned—its first two layers remain frozen while the third and fourth blocks are adapted to OCT textures—and a classical projection reduces the 512-dimensional feature vector to a 5-dimensional input for a VQC acting on five qubits. Hyperparameters such as qubit count and circuit depth are selected via Particle Swarm Optimization (PSO). Experiments on the OCT2017 dataset show that the proposed model achieves 89.38% accuracy, supporting the feasibility of a parameter-efficient hybrid alternative for medical image classification under NISQ-inspired constraints. The results support the feasibility of hybrid quantum–classical models for medical imaging in the Noisy Intermediate-Scale Quantum era.

**Keywords:** Quantum Transfer Learning · Medical Image Classification · Optical Coherence Tomography · Hybrid Neural Networks · Particle Swarm Optimization

## 1 Introduction

Optical Coherence Tomography (OCT) produces high-resolution cross-sectional retinal images and is widely used in clinics for early diagnosis and monitoring of retinal diseases. Manual interpretation of OCT volumes is time-consuming and

subject to inter-observer variability, motivating automatic classification systems. Convolutional neural networks (CNNs) and transfer learning achieve state-of-the-art results on medical images[1–3], yet their large parameter counts and training cost may limit deployment. Recent advances in quantum computing and quantum machine learning suggest that hybrid quantum–classical models could offer parameter-efficient alternatives[4–6].

As emphasised in ophthalmology literature, the early detection of retinal pathologies such as choroidal neovascularisation, diabetic macular oedema and drusen is vital for preventing vision loss. OCT provides micrometre-resolution cross-sections of retinal layers, allowing clinicians to visualise subtle structural changes. However, analysing large OCT volumes manually places a heavy workload on experts and may introduce subjective bias. Automated interpretation based on deep learning seeks to assist clinicians by rapidly screening for disease patterns with high sensitivity and specificity, while maintaining reproducibility across hospitals.

Traditional CNNs excel at feature extraction but typically involve tens of millions of learnable parameters and require significant GPU resources for training and inference. Such computational demands may limit their adoption in clinical settings where hardware is scarce or rapid deployment is needed. Moreover, purely classical models are sometimes criticised for limited flexibility in modelling decision boundaries under class imbalance or small dataset regimes. These constraints motivate the exploration of hybrid quantum–classical strategies that aim to reduce the number of trainable parameters while preserving classification accuracy.

In parallel, the field of quantum computing has witnessed rapid progress, yet current devices operate in the Noisy Intermediate-Scale Quantum (NISQ) regime with limited qubits and coherence times. Hybrid approaches that offload heavy feature extraction to classical networks and use shallow quantum circuits for decision making represent a feasible pathway towards near-term quantum advantage. In this context, Quantum Transfer Learning (QTL) leverages pretrained classical encoders to generate compact feature vectors that are subsequently processed by variational quantum circuits (VQCs). The integration of classical and quantum modules yields a learning pipeline that is both hardware-aware and adaptable to real-world medical imaging tasks.

Hybrid schemes leverage classical networks for feature extraction while delegating the decision stage to a variational quantum circuit (VQC). In Quantum Transfer Learning (QTL), a pretrained CNN provides high-level feature vectors that are encoded into a quantum state and processed by a parametrised quantum circuit[7, 8]. The circuit parameters are updated via classical optimisation, reflecting the variational quantum algorithm approach[9]. Current NISQ devices have limited qubit counts and are susceptible to noise; therefore, hybrid architectures offer a feasible intermediate solution between model expressivity and hardware limitations.

This paper proposes a QTL approach tailored for multi-class OCT classification. We partially fine-tune a ResNet34 backbone on the OCT2017 dataset,

project its 512-dimensional feature vectors onto a five-dimensional subspace, and classify them with a depth-12 VQC acting on five qubits. Particle Swarm Optimization (PSO) selects optimal values for qubit count, circuit depth and other hyperparameters. The contributions are: (i) demonstrating a hybrid CNN–VQC architecture with partial fine-tuning for OCT classification, (ii) introducing PSO-based hyperparameter selection for quantum circuits, and (iii) reporting experimental results that evaluate the feasibility of hybrid classification under constrained quantum resources.

The remainder of the paper is organised as follows. Section 2 summarizes related hybrid quantum–classical approaches. Section 3 describes the pretrained CNN, VQC design, PSO optimisation and the full QTL pipeline. Section 4 details the dataset, evaluation metrics and experimental setup. Section 5 presents the results and comparisons. Conclusions and future research directions are discussed in Section 6.

## 2 Related Work

Hybrid quantum–classical learning combines classical representation learning with quantum decision making to address the limitations of NISQ devices. Early studies explored variational quantum classifiers operating on small classical feature sets[5]. Matic et al. evaluated several hybrid CNN–quantum classifiers for radiology and observed performance comparable to classical networks under controlled settings[10]. Azevedo et al. applied transfer learning with quantum layers to mammography, reporting improved generalisation[11]. Khan et al. proposed a hybrid quantum CNN for brain tumour classification, demonstrating high accuracy with fewer parameters[12]. Radhi et al. reviewed quantum machine learning in medical imaging and emphasised the need for standardised datasets and benchmarking protocols[13].

Subsequent works refined the quantum layer design. Ajlouni et al. introduced adaptive hybrid quantum CNNs with tunable entanglement structures for various imaging tasks[14]. Senokosov et al. compared parallel quantum circuits and quanvolution layers on MNIST and Medical MNIST, highlighting the potential of quantum feature mapping[15]. Jain et al. applied QTL to fundus images for diabetic retinopathy detection, providing preliminary evidence of quantum advantages on retinal data. While these studies illustrate promising directions, systematic evaluations on multi-class OCT classification remain limited. Our work addresses this gap by integrating a partially fine-tuned ResNet34 encoder with a VQC and systematically tuning quantum hyperparameters via PSO.

Beyond these pioneering efforts, several authors have examined how classical transfer learning can be seamlessly combined with variational quantum circuits. For instance, the work by Radhi et al. synthesises results from a broad range of applications—ranging from dermatological lesion analysis to volumetric MRI classification—and concludes that hybrid architectures often match classical accuracy with far fewer trainable weights. Ajlouni et al.'s adaptive quantum CNN explores dynamic entanglement patterns to adapt the quantum layer's expressive

power to the complexity of the input data. Senokosov et al. demonstrate that even simple quanvolutional layers, which apply fixed quantum filters to image patches, can enrich feature representations when appended to classical CNNs. These studies collectively highlight that quantum circuits are versatile components that can be tailored to specific imaging modalities and that their integration into classical frameworks requires careful hyperparameter tuning. Despite these advances, the application of QTL to OCT data—where three disease classes and normal retina must be discriminated—remains largely unexplored. Therefore, our study contributes by systematically evaluating a hybrid CNN–VQC model on the OCT2017 benchmark with a rigorous PSO-based hyperparameter search.

## 3 Methodology

Our hybrid framework comprises a classical encoder, a variational quantum classifier and a PSO-driven hyperparameter search. Key components are summarised below.

### 3.1 Pretrained CNN and Feature Extraction

We employ a ResNet34 pretrained on ImageNet as the classical encoder. To adapt high-level representations to OCT data while maintaining general features, we freeze the initial two residual blocks and fine-tune the third and fourth blocks on the OCT2017 training set. For an input image $\mathbf{x}$, the encoder outputs a 512-dimensional feature vector $\mathbf{h} = \phi(\mathbf{x}; \psi)$ defined by Eq. (1). The network's class probabilities are obtained using the softmax function in Eq. (2).

$$\mathbf{h} = \phi(\mathbf{x}; \psi), \quad \mathbf{h} \in \mathbb{R}^{512} \tag{1}$$

$$\hat{\mathbf{p}} = \mathrm{softmax}(\mathbf{o}), \quad \hat{p}_c = \frac{\exp(o_c)}{\sum_{j=1}^{C} \exp(o_j)} \tag{2}$$

ResNet34 comprises an initial convolutional stem followed by four sequential residual blocks that facilitate the training of deep networks by introducing identity skip connections[2]. These residual connections allow gradients to flow directly to earlier layers, mitigating the vanishing gradient problem. In transfer learning scenarios, the early convolutional layers tend to extract generic edge and texture primitives, whereas deeper layers learn task-specific abstractions. By freezing the first two blocks, we preserve features learned from the large-scale ImageNet dataset, which are also relevant for OCT textures. Fine-tuning the third and fourth blocks allows the model to specialise its representations to the subtle morphological differences between CNV, DME, drusen and normal retina while keeping the parameter count modest.

### 3.2 Classical–Quantum Projection and VQC

The high-dimensional features are projected to a low-dimensional vector $\mathbf{z} = g(\mathbf{h};\omega)$ (Eq. (3)) using a learnable linear layer, reducing the 512 dimensions to $n$ elements matching the number of qubits. Each component of $\mathbf{z}$ determines the rotation angle of a single-qubit gate. The variational quantum circuit, specified in Eq. (4), consists of angle embedding followed by layers of parameterised single-qubit rotations and entangling CNOT gates. The measurement of $n$ Pauli-$Z$ expectation values yields a real-valued vector $\mathbf{s}(\mathbf{z}) \in \mathbb{R}^n$. These expectations are mapped via a classical post-processing network to the final output.

$$\mathbf{z} = g(\mathbf{h};\omega), \quad \mathbf{z} \in \mathbb{R}^n \tag{3}$$

$$s_k(\mathbf{z}) = \left\langle 0^{\otimes n} \middle| U^\dagger(\mathbf{z},\theta)\, Z_k\, U(\mathbf{z},\theta) \middle| 0^{\otimes n} \right\rangle, \quad k = 1,\ldots,n \tag{4}$$

We adopt an angle-embedding strategy wherein each element of $\mathbf{z}$ is encoded into the rotation angle of a $R_y$ gate acting on an individual qubit[5]. This encoding preserves the ordering of features and enables continuous parameterisation of qubit states. Following the encoding layer, we stack $d$ layers of a strongly entangling ansatz composed of trainable $R_y$ rotations and entangling CNOT gates arranged in a circular pattern. Entanglement is essential for modelling correlations that cannot be captured by separable qubit states, and the circuit depth $d$ controls the expressive power of the quantum model. In our experiments, PSO selects $n = 5$ and $d = 12$ as optimal settings. After measurement, the resulting expectation values are passed to a classical dense layer with $h$ hidden units (here $h = 166$) and a ReLU activation, followed by a softmax output. This post-processing stage allows the hybrid model to learn complex decision boundaries in classical space while keeping the quantum circuit shallow.

### 3.3 Particle Swarm Optimisation

Selecting the qubit count $n$, circuit depth $d$ and learning rate $\eta$ crucially impacts performance. We adopt Particle Swarm Optimisation (PSO) for data-driven hyperparameter tuning. Each particle encodes a candidate configuration $[n, d, \eta, q_\delta, h]$, where $q_\delta$ is a circuit initialisation scale and $h$ denotes the number of hidden units in the post-processing network. Particle positions and velocities are updated according to Eqs. (5)–(6). The objective function, defined in Eq. 7, maximises validation accuracy over a small number of epochs. Table 1 lists the optimal hyperparameters obtained by PSO and used in final training.

$$\mathbf{v}_i^{t+1} = \omega\, \mathbf{v}_i^t + c_1 r_1(\mathbf{p}_i^t - \mathbf{x}_i^t) + c_2 r_2(\mathbf{g}^t - \mathbf{x}_i^t), \tag{5}$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \tag{6}$$

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}\in\Omega}\ \mathrm{Acc}_{\mathrm{val}}(\boldsymbol{\lambda}) \tag{7}$$

**Table 1.** Best hyperparameters obtained via PSO (configuration used in the final training).

| **Hyperparameter** | **Value** |
|---|---|
| Learning rate | 0.002587 |
| Quantum circuit depth ($q_{\text{depth}}$) | 12 |
| Quantum step parameter ($q_\delta$) | 0.0286 |
| Number of hidden units | 166 |
| Batch size | 60 |
| Number of qubits ($n_{\text{qubits}}$) | 5 |

PSO is a population-based heuristic that simulates a flock of particles exploring the search space[16]. At each iteration, a particle moves towards a weighted combination of its own best-known position and the swarm's global best. The inertia term encourages exploration, while cognitive and social coefficients balance individual and collective learning. Unlike gradient-based methods, PSO does not require derivative information and can handle discrete or non-smooth objective landscapes. In our implementation, we initialise the swarm with random hyperparameter vectors drawn from a predefined range, evaluate the corresponding models on the validation set using a few training epochs, and iteratively update particle positions. After convergence, the best configuration is adopted for full training on the combined training and validation sets.

### 3.4 Quantum Transfer Learning (QTL)

Quantum Transfer Learning (QTL) extends the classical transfer learning approach to the domain of quantum machine learning by combining a pretrained classical model with a parametrised quantum circuit. In classical transfer learning, features learned on a source task are reused on a target task; QTL retains this idea but performs the final decision stage on a quantum device. The aim is to exploit the strong representational power of deep neural networks while capitalising on the expressive capacity of quantum circuits for low-dimensional decision boundaries. Heavy data processing and feature extraction remain on the classical side, whereas the quantum circuit handles the classification of a compact feature vector. This division of labour is particularly suitable for current NISQ devices, which offer a limited number of qubits and are sensitive to noise[17, 6].

In the QTL framework, the pretrained classical encoder is typically used as a fixed or partially fine-tuned feature extractor, preserving the knowledge acquired from large datasets and avoiding the need to train millions of parameters in tandem with the quantum circuit. The quantum component, usually a variational quantum circuit, serves as the learnable classifier. Its parameters are updated using classical optimisation routines, leading to a training loop closely resembling that of classical neural networks[5, 4]. This hybrid training scheme enables the seamless integration of quantum models into established deep learning workflows.

**Concept and Definition** The core idea behind QTL is to leverage a pretrained classical network $\phi$ to map an input image $\mathbf{x}$ to a high-dimensional feature vector $\mathbf{h} = \phi(\mathbf{x}; \psi)$ and then use a quantum circuit as a classifier acting on a low-dimensional representation of $\mathbf{h}$. The classical network provides rich, transferable representations learned from large image datasets, while the quantum circuit offers a potentially more expressive decision function with a small number of trainable parameters. Because the classical features are reused, the quantum circuit can remain shallow—an important consideration for NISQ hardware. During training, only the parameters of the projection layer, the quantum circuit and the post-processing network are updated, keeping the early convolutional layers of the encoder fixed. This approach reduces the trainable parameter burden of the decision stage and may help mitigate overfitting when labelled medical data are scarce.

**Classical-to-Quantum Feature Mapping** A critical step in QTL is mapping the high-dimensional classical features $\mathbf{h} \in \mathbb{R}^{512}$ to a vector $\mathbf{z} \in \mathbb{R}^n$ compatible with an $n$-qubit quantum circuit. In our framework, this is accomplished by a learnable linear projection $g$ as defined in Eq. (3). The projection reduces the dimensionality of the feature space from 512 to $n$ (with $n = 5$ in our experiments) and is trained jointly with the quantum circuit. Although such a drastic reduction may induce information loss, the projection layer and the variational quantum circuit are trained jointly so that the compressed representation can still retain discriminative information relevant to the final classification task. After projection, each element of $\mathbf{z}$ is encoded into the quantum circuit via angle embedding, where the value determines the rotation angle of a single-qubit $R_y$ gate. This embedding procedure allows continuous classical values to influence quantum states directly. The choice of projection dimension $n$ and the specific encoding strategy are design decisions that influence the balance between hardware constraints and model capacity[5].

The classical–quantum feature mapping thus functions as a bridge between the extensive feature representations learned by the CNN and the restricted qubit space of the quantum circuit. By optimising the projection layer and the quantum circuit parameters together, the QTL model adapts the classical features to the quantum domain while supporting effective decision formation through subsequent classical post-processing.

### 3.5 Quantum Transfer Learning Pipeline

Figure 1 illustrates the complete pipeline. An OCT image is passed through the partially fine-tuned ResNet34 to extract 512 features. These are linearly projected to a 5-dimensional vector (for $n = 5$ qubits) and encoded into the quantum circuit via angle embedding. The depth-12 VQC produces expectation values which are fed into a classical network with 166 hidden units and a softmax output. The entire projection layer, quantum circuit parameters and post-processing weights are trained end-to-end using the cross-entropy loss on

the training set. Hyperparameters are fixed from the PSO stage to ensure reproducibility.

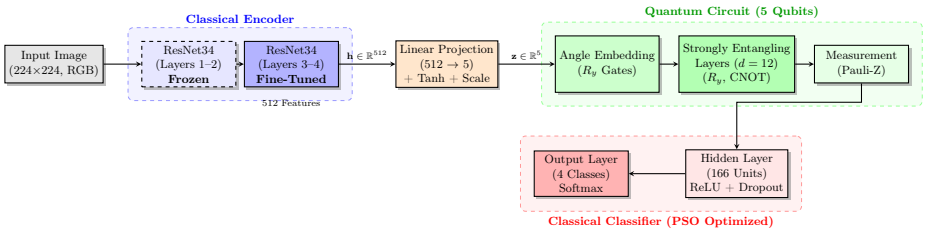



**Fig. 1.** The proposed hybrid architecture. The ResNet34 backbone is partially fine-tuned (Layers 3–4) to extract 512-dimensional features. These are projected to a 5-qubit quantum state, processed by a variational quantum circuit (depth=12), and classified by a classical dense network (166 hidden units) optimised via PSO.

## 4 Experiments

### 4.1 Dataset and Preprocessing

We evaluate the proposed QTL framework on the OCT2017 dataset, which contains OCT slices of four classes: CNV, DME, drusen and normal retina[3]. Images are resized to $224 \times 224$ pixels and converted to three-channel inputs by duplicating the grayscale image across RGB channels. To enhance generalisation, random rotations and horizontal flips are applied during training. We build a balanced subset with 200 images per class for training and validation, reserving a separate test set for performance assessment.

The OCT2017 dataset comprises over 109,000 B-scan slices drawn from 4,686 patients, with class distributions heavily skewed towards CNV and normal categories[3]. To manage the computational cost of quantum simulation while ensuring class balance, we sample 800 images (200 per class) for training and validation and another 800 images for testing. Prior to augmentation, each grayscale image is histogram normalised and padded to preserve aspect ratio before resizing to the required input dimensions. During training, we apply random 10-degree rotations and horizontal flips with probability 0.5 to mitigate overfitting. No augmentation is applied to the validation or test sets.

### 4.2 Experimental Setup

Training is performed on the TRUBA high-performance computing infrastructure using CUDA-accelerated GPUs for the classical network and the simulation of the quantum circuit. The quantum circuit is simulated under a noise-free assumption. Training and validation splits are fixed, and PSO uses only the validation subset. The performance metrics—accuracy, precision, recall and F1-score—are

computed on the held-out test set using Eqs. (8)–(9). A confusion matrix visualises class-wise predictions.

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} \text{TP}_i}{N} \tag{8}$$

$$\text{F1}_{\text{macro}} = \frac{1}{C} \sum_{i=1}^{C} \frac{2 \cdot \text{TP}_i}{2 \cdot \text{TP}_i + \text{FP}_i + \text{FN}_i} \tag{9}$$

Our experiments utilise the PyTorch framework for classical computations and PennyLane for quantum circuit simulation. Each model is trained using the Adam optimiser with the learning rate $\eta$ selected by PSO and a mini-batch size of 60. Early stopping on the validation accuracy prevents overfitting. The classical encoder's fine-tuned layers and the projection, quantum circuit and post-processing parameters are trained jointly for 30 epochs using cross-entropy loss. Quantum simulations are performed in an idealised noise-free environment; investigating the effect of realistic noise models and error-mitigation techniques is left for future work.

## 5 Results

This section presents the experimental results of the proposed Quantum Transfer Learning (QTL) approach on the OCT2017 dataset. All results were obtained on the test dataset that was kept completely independent from the training and hyperparameter optimization processes. The performance metrics of the final experiments performed using the optimal configuration determined by PSO-based hyperparameter optimization are reported below.

The proposed QTL model achieved 89.38% accuracy on the test set. This result shows that the hybrid classical–quantum approach provides a feasible and stable performance under the evaluated experimental setting for the multi-class medical image classification problem. The macro-averaged precision, recall and F1-score were calculated as 0.8948, 0.8938 and 0.8929, respectively. These metrics indicate that the model exhibits a balanced performance across classes and does not overfit to a single class.

When the class-wise performance analysis is examined, it is seen that the highest F1-score (0.9417) is obtained for the NORMAL class. This indicates that the model can distinguish healthy retina images with high accuracy. The F1-scores for the CNV and DME classes were calculated as 0.8822 and 0.8889, respectively, indicating high and consistent classification performance for these classes as well. In the DRUSEN class, however, the recall is relatively lower (0.8050), while the precision is high (0.9200). This result shows that the model makes more cautious decisions in the DRUSEN class and limits false positive predictions.

The class-wise precision, recall and F1-score values obtained are presented in Table 2. The results show that the model generally exhibits balanced performance across classes; however, the DRUSEN class has relatively low recall.

**Table 2.** Class-wise performance results on the test set.

| Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| CNV | 0.8844 | 0.8800 | 0.8822 | 200 |
| DME | 0.8598 | 0.9200 | 0.8889 | 200 |
| DRUSEN | 0.9200 | 0.8050 | 0.8587 | 200 |
| NORMAL | 0.9151 | 0.9700 | 0.9417 | 200 |
| Accuracy | | | 0.8938 | 800 |
| Macro avg | 0.8948 | 0.8938 | 0.8929 | 800 |
| Weighted avg | 0.8948 | 0.8938 | 0.8929 | 800 |

**Table 3.** Comparison of the proposed hybrid QTL approach with state-of-the-art studies.

| Study | Method | Accuracy (%) |
|---|---|---|
| Kermany et al. [3] | InceptionV3 (Transfer Learning) | 96.60 |
| Lee et al. [18] | VGG16 (Deep Learning) | 93.45 |
| Rong et al. [19] | Surrogate-assisted CNN | 94.20 |
| **Proposed Method** | **Hybrid ResNet34 + VQC (Ours)** | **89.38** |

The class-wise decision behaviour of the model was examined in more detail via the confusion matrix presented in Fig. 2. A closer inspection shows that the NORMAL class is the most separable category, with the lowest error rate among all classes. The DME and CNV classes also exhibit strong classification performance, although a limited degree of confusion is observed between them, which may be attributed to partially overlapping retinal patterns. In contrast, DRUSEN remains the most challenging class, consistent with its lower recall reported in Table 2. Overall, the confusion structure indicates that most errors are concentrated among clinically similar disease categories, whereas healthy retina images remain comparatively easier to distinguish.

The PSO-based hyperparameter optimization determined that 5 qubits, a quantum circuit depth of 12, and a learning rate of approximately 0.0026 provide a balanced structure between the expressive power of the variational quantum circuit and training stability. The results obtained under this configuration show that the quantum classifier component functions effectively as a decision maker on the classical features extracted by the pretrained CNN.

As seen in Table 3, classical deep learning models still achieve higher classification accuracy on OCT2017. Nevertheless, the proposed hybrid model operates under a highly compressed 512-to-5 feature interface and a limited quantum parameter space. Therefore, its contribution should be interpreted as a parameter-efficient proof-of-concept for low-dimensional hybrid decision modeling under constrained quantum resources, rather than as evidence of superiority over high-capacity classical CNNs. The results indicate that a fine-tuned ResNet34 backbone coupled with a variational quantum circuit can preserve discriminative

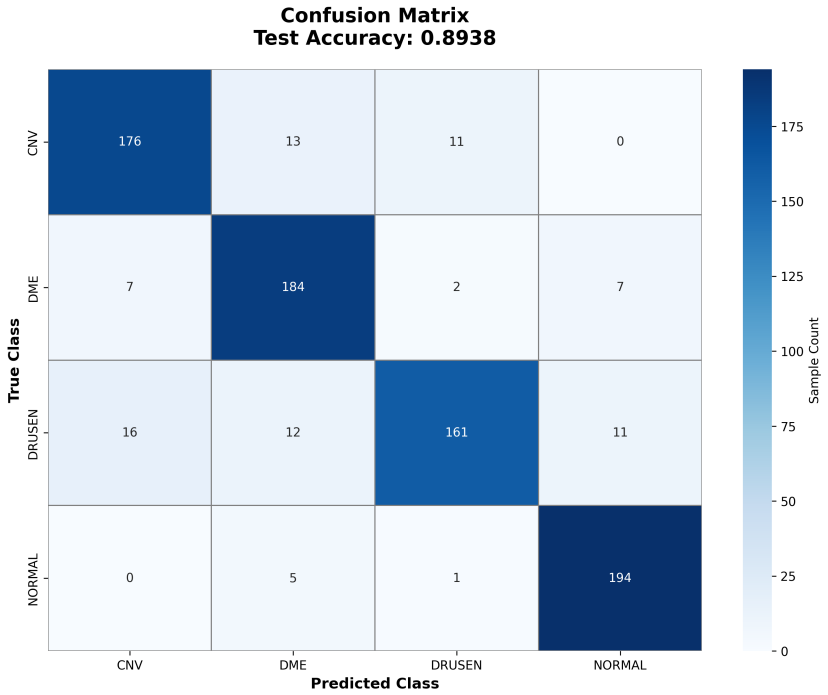



**Fig. 2.** Confusion matrix of the proposed QTL model on the independent test set. Each class contains 200 samples, enabling direct inspection of class-wise error distributions and absolute misclassification counts.

class information under severe dimensional and quantum resource constraints. In this respect, the study supports the feasibility of quantum transfer learning for medical imaging in the NISQ-inspired setting considered here.

Overall, the experimental results obtained show that the proposed Quantum Transfer Learning approach provides a feasible, balanced and interpretable performance for the multi-class medical image classification problem. When the quantitative performance metrics and confusion matrix analysis are evaluated together, it can be seen that the hybrid classical–quantum architecture makes a meaningful contribution to the classification process.

## 6 Conclusion

In this study, a hybrid Quantum Transfer Learning (QTL) approach that brings together classical and quantum learning components for the multi-class medical image classification problem has been proposed. The proposed method aims to leverage the powerful representation extraction capacity of a pretrained convolutional neural network and combine it with a variational quantum circuit, in order to exploit quantum learning models at the decision stage.

Experimental results, based on evaluations performed on the OCT2017 dataset, show that the proposed QTL approach provides a feasible and stable proof-of-concept for the multi-class medical image classification problem. The results obtained under the quantum circuit configuration determined by PSO-based hyperparameter optimization indicate that the quantum classifier component can operate effectively on the deep features extracted by the classical CNN within the evaluated experimental setting. Class-wise performance metrics and confusion matrix analysis further show that the model can draw balanced distinctions among classes, while most errors remain concentrated among clinically similar categories.

An important advantage of the proposed approach is the reduction of the parameter count by freezing the first two layers of the ResNet34 backbone and fine-tuning the deeper layers (Layers 3–4), while allowing the quantum circuit to be trained in a smaller, controllable parameter space. This structure provides a useful framework for studying hybrid quantum–classical models under near-term quantum constraints, particularly in settings where parameter efficiency and low-dimensional decision modeling are of interest. Moreover, PSO-based hyperparameter selection allowed for a more systematic model configuration by reducing experimental bias in quantum circuit design.

However, in this study, quantum circuit simulations were conducted under the assumption of an ideal and noise-free environment. The effects of noise and errors that may arise on real quantum hardware were left outside the scope of this work. Future studies plan to evaluate the proposed QTL approach on noisy quantum circuits, to investigate different quantum ansatz structures, and to conduct comprehensive comparisons with larger datasets. In addition, the applicability of the quantum component to different clinical imaging modalities is regarded as an important direction to be investigated in the future.

In conclusion, this study demonstrates that hybrid approaches based on the combined use of classical deep learning and quantum learning models offer a promising research direction for medical image classification problems. The findings support that the Quantum Transfer Learning approach provides a viable and open framework for research in the context of near-term quantum technologies.

**Acknowledgments.** This research was supported by the TUBITAK 2211/C National PhD Scholarship Program in the Priority Fields in Science and Technology entitled "Improving the Performance of Machine Learning Algorithms with Quantum Computing".

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

## References

1. Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
3. Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018.
4. Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
5. Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.
6. John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.
7. Andrea Mari, Thomas R Bromley, Josh Izaac, Maria Schuld, and Nathan Killoran. Transfer learning in hybrid classical-quantum neural networks. *Quantum*, 4:340, 2020.
8. Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.
9. Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016.
10. A. Matic, Maureen Monnet, J. Lorenz, Balthasar Schachtner, and Thomas Messerer. Quantum-classical convolutional neural networks in radiological image classification. pages 56–66, 09 2022.
11. Vanda Azevedo, Carla Silva, and Inês Dutra. Quantum transfer learning for breast cancer detection. *Quantum Machine Intelligence*, 4(1):5, 2022.
12. Muhammad Al-Zafar Khan, Abdullah Al Omar Galib, Nouhaila Innan, and Mohamed Bennai. Brain tumor diagnosis using quantum convolutional neural networks. *arXiv preprint arXiv:2401.15804*, 2024.
13. Eman A Radhi, Mohammed Y Kamil, and Mazin Abed Mohammed. Quantum machine and deep learning for medical image classification: A systematic review of trends, methodologies, and future directions. *Iraqi Journal for Computer Science and Mathematics*, 6(2):9, 2025.
14. Naim Ajlouni, Adem Özyavaş, Mustafa Takaoğlu, Faruk Takaoğlu, and Firas Ajlouni. Medical image diagnosis based on adaptive hybrid quantum cnn. *Research Square (Research Square)*, 06 2023.
15. Arsenii Senokosov, Alexandr Sedykh, Asel Sagingalieva, Basil Kyriacou, and Alexey Melnikov. Quantum machine learning for image classification. *Machine Learning Science and Technology*, 5:15040–15040, 02 2024.
16. James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. ieee, 1995.
17. Maria Schuld and Francesco Petruccione. Supervised learning with quantum computers. *Quantum science and technology*, 17, 2018.
18. Cecilia S Lee, Doug M Baughman, and Aaron Y Lee. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. *Ophthalmology Retina*, 1(4):322–327, 2017.

19. Yibiao Rong, Dehui Xiang, Weifang Zhu, Kai Yu, Fei Shi, Zhun Fan, and Xinjian Chen. Surrogate-assisted retinal oct image classification based on convolutional neural networks. *IEEE journal of biomedical and health informatics*, 23(1):253–263, 2018.

# System Analysis of Acne Diagnostics Using Artificial Intelligence

Shavkat Fazilov[1][0000-0002-4217-6659] and Shukhrat Mamadjanov[1,2][0009-0001-3036-6607]

[1] Digital Technologies and Artificial Intelligence Development Research Institute, Tashkent, Uzbekistan
[2] Tashkent International University, Tashkent, Uzbekistan
shukhratjonmamadjanov@gmail.com

**Abstract.** Acne, considered one of the dermatological disorders, is a widespread skin disease that often occurs during crucial developmental stages in adolescents and can lead to long-term psychosocial consequences. Traditional methods, such as assessing overall severity and counting rash elements, are limited by subjectivity and time constraints. This article focuses on a systematic review of the latest advancements in diagnosing acne, detecting and counting rash elements, and assessing severity using artificial intelligence (AI), highlighting the potential of AI-based methods to improve objectivity, reproducibility, and clinical effectiveness. A comprehensive literature search was conducted in PubMed, Scopus, arXiv, Embase, and Web of Science databases for studies published from 2017 to February 2026. The search strategy included terms related to "acne" and various AI methodologies (e.g., "neural network," "deep learning," "convolutional neural network"). During the research process, 391 articles were reviewed, of which 39 articles met the final criteria. Data were extracted on study design, dataset characteristics (including internal and open-access resources such as ACNE04 and AcneSCU), AI architectures (primarily CNN-based models), and performance indicators. Although AI-based models have demonstrated high accuracy under controlled conditions, the scarcity of large public datasets, the predominance of data from specific ethnic groups, and the lack of comprehensive external validation reveal significant barriers to implementation in clinical practice. The research findings indicate that while AI has the potential to standardize acne assessment, reduce observer variability, and enable self-monitoring through mobile platforms, there are considerable challenges in achieving reliable real-world application. Future research should prioritize the creation of large, diverse, and openly accessible datasets and conduct prospective clinical trials to ensure fair and effective dermatological care.

**Keywords:** Acne, artificial intelligence, acne position, dermatology, image classification, computer vision.

## 1 Introduction

Facial acne is one of the most common skin problems, especially prevalent among adolescent boys and girls, and is a condition that affects a large portion of the younger

generation [1]. This ailment typically emerges during youth, at a crucial stage of a person's social, psychological, and physical development, potentially leading to dissatisfaction with one's appearance and negatively impacting social relationships and self-confidence. Furthermore, in many cases, this condition persists beyond puberty, resulting in a decrease in the quality of life [2].

Accurate measurement and assessment of acne is crucial for effective treatment and the reproducibility of scientific research [2]. However, due to the spontaneous variability of the acne and the uneven distribution of the rash, this process can be complicated. In another study, a new hybrid model was presented, combining DeepLabV3 and InceptionV3 for classification, which offers an improved solution for detecting acne. The DeepLabV3 model separates acne lesions and creates distinct segmentation masks, while InceptionV3 effectively classifies various types of acne and increases overall diagnostic accuracy [3].

Traditional methods used to determine the degree of acne are divided into two main areas: general assessment and counting of inflammatory elements. However, both of these methods are limited and time-consuming, as they are based on personal considerations[4]. Subjective assessment of severity on a general scale is carried out by comparing the clinical picture with a standard picture or written description. Currently, the most commonly used scale for assessing acne is the Investigator Global Assessment (IGA), which ranges from 0 to 4 (0, pure; 1, almost transparent; 2, light; 3, average; 4, severe) [4] and the Global Acne Severity Scale (GEA) (0, no lesions; 1, almost no injuries; 2, light; 3, average; 4, heavy; 5, very heavy) [4].

Another possible approach is the method of counting lesions, shown in the example of the Hayashi scale. This method assesses the severity of the acne by counting inflammatory lesions (papules and pustules) on one side of the face and divides it into four categories: mild (0-5 lesions), moderate (6-20), severe (21-50) and very severe (>50) [4]. Assessment of the overall severity is usually simpler and faster than counting lesions, but this method depends more on the evaluator and is less sensitive to changes in the disease over time. On the contrary, counting lesions provides a more accurate and reproducible method, but it is often impractical, time-consuming, and does not take into account such parameters as the nature of the lesion's spread, size, or the presence of erythema[4]. The criteria for counting lesions and assessing the severity of acne can be further divided into two types: criteria that take into account only the number of lesions, and criteria that take into account both the number and type of lesions (e.g., open and closed comedones, papules, pustules, and nodules) [4].

Recently, thanks to the rapid development of artificial intelligence (AI) and computer vision technologies, new opportunities have emerged for automatic and objective assessment and detection of acne. This process is carried out, in particular, by segmenting the affected areas and calculating their quantity, as well as by assessing the overall situation. For the purpose of high-precision detection and grouping of acne symptoms, models based on artificial intelligence, prepared using vast amounts of data and advanced computational methods, allow for the assessment of this skin problem and the effectiveness of treatments performed in relation to it using standardized and continuously repeatable methods [5]. The use of automated methods for assessing acne based

on the principles of artificial intelligence allows for faster results compared to conventional assessment methods, ensures high accuracy, increases recovery, and normalizes the assessment process. This serves to minimize differences (variations) in assessment between different observers and within one observer [6]. AI algorithms can also lead to self-monitoring through mobile applications, evaluating treatment responses, processing large volumes of epidemiological data[7], and simplifying comparisons between studies in meta-analysis[8].
The purpose of this systematic review is to identify acne, quantify their prevalence, and comprehensively cover the latest developments and applications of artificial intelligence in assessing the severity of the disease.

## 2 Materials and methods

A systematic review of the literature included a thorough examination of the analyzed articles, paying particular attention to the structure of the computational models used (e.g., CNNs, deep residual transformations, and integrated systems), the nature of the data sources, the methods used for marking/tagging, and the procedures for evaluating the results. The comparative aspect focuses on key performance indicators, including accuracy, retention, true negative score, F1 score, and AUC score. All of this process strictly adhered to the guidelines set out in the PRISMA 2020 documents to guarantee accuracy in methodology, reproducibility, and thoroughness in selected studies. A comprehensive search was carried out in the databases PubMed, Scopus, Web of Science, IEEE Xplore, arXiv, Embase. A methodological search plan was compiled for searching relevant literature, which included designated headings and general linguistic terms for the subject. The resulting survey combined acne-related terms with artificial intelligence and visual data evaluation phrases, using logical connectors such as AND and OR. In the process of selecting preliminary results, comprehensive reviews or studies combined with meta-analysis were deliberately skipped, which ensured the emphasis on basic, live, or laboratory research in the final selection. The study included works published in the period from January 2017 to February 2026. We included in this analysis any study covering patients of all ages with a confirmed presence of acne and using artificial intelligence or computer vision algorithms to identify the damage or assess its severity. Studies that yielded good results had to compare their models with clinical data based on "correct" standards evaluated by doctors (dermatologists) or approved by specialists. At the same time, they had to show at least one numerical performance metric (e.g., accuracy level, Dice coefficient, sensitivity, or specificity). The study examined observational studies published in English. We excluded conference summaries, non-English-language articles, descriptive reviews and meta-analyses, as well as any research that reused data sets already used elsewhere to avoid duplication.

# 3 Results

## 3.1 Sorting research and transparency

As a result of the systematic search, 391 articles were identified, of which 115 studies underwent a full-text review. As a result, 39 studies were selected. The analysis was focused on the methods of artificial intelligence used in the diagnosis and assessment of the severity of acne, the data sets used, and the results achieved. Of these, 17 (43.6%) were based solely on internal datasets, while 22 (56.4%) used publicly available datasets (Face Datasets [3], Dermnet [9], ACNE04 [11], AcneSCU [12], MedDialog [13]). More specifically, four studies (10.2%) used only the ACNE04 dataset, two studies (5%) used the AcneSCU dataset, nine studies (23%) combined ACNE04 with internal data, one study (2.5%) used the Face Datasets dataset, two studies (5%) used the Dermnet dataset, two studies (5%) used the MedDialog dataset, and three studies (7.6%) combined ACNE04, AcneSCU, and internal data.
In addition, in one study, a synthetic data set created by artificial intelligence trained using a Generative Adversarial Network (GAN) was used in the ACNE04 dataset [14]. In another study[14], an open data set of faces (2037 images from CelebAMask-HQ and Flickr-Faces-HQ) was used, a small portion of which suffered from acne. It should be noted that in most studies, images from outpatient clinics were used, but in seven studies (28%) images taken by the patients themselves using smartphones were used. Publicly available software code was available in ten studies (25.6%).

## 3.2 Demographic characteristics of participants and diversity of the dataset

The age of the participants was recorded in only five studies (12.8%). Data on the ethnicity of the participants were available for 16 studies (41%), including those who used only the open ACNE04 database. Of these, 15 (38.4%) mainly included persons of East Asian or Chinese nationality. The only exception was a study by Seitè et al., which analyzed 5,972 images of 1,072 acne patients representing various ethnic groups, such as Europeans, Africans, Asians, Latin Americans, and Indians, in countries such as France, South Africa, China, and India. Images were obtained using both iOS and Android smartphones, which ensured a wide range of skin colors and types.

## 3.3 Algorithms

In 13 (33.3%) of the analyzed 39 studies, multiple deep learning (DL) algorithms were used. CNN and deep convolutional neural networks (DCNN) were the main architectures, accounting for 72% (28/39) of all studied models (Table 1). The main focus of the study was the segmentation and counting of acne (17 studies, 43.6%) and determining the severity (20 studies, 51.2%), while only two studies (5.2%) studied the binary classification (presence and absence of acne). The training datasets varied significantly, ranging from 276 to over 212,374 images. Key methodological approaches include CNN-based architectures, semi-controlled teaching methods, and generative competitive networks (GANs) such as StyleGAN2-ADA. These were used to create synthetic datasets, thereby eliminating data shortages and increasing the generalizability of models. The training data was typically divided into training and validation sub-sets, with 80-90% directed to training and the remainder to validation.

**Table 1.** Description of the works included in the study.

| Authors | Application Category | Methods | Dataset | Dataset Source | Publicly Available Code | Number of Photos |
|---|---|---|---|---|---|---|
| Sharmin et al. [3] (2025) | A Hybrid CNN Framework DLI-Net for Acne Detection with XAI | Introduces a novel hybrid model that combines DeepLabV3 for precise image segmentation with InceptionV3 for classification, offering an enhanced solution for acne detection. | Face Datasets consists of 1725 images of inflammatory acne, 594 images of non-inflammatory acne, and 8460 images of clear skin. | To address the class imbalance, oversampling techniques such as data augmentation (rotation and flipping) were applied, resulting in a balanced dataset with 5051 images of inflammatory acne, 1962 images of non-inflammatory acne, and 5325 images of clear skin. | Datasets available at: https://www.kaggle.com/datasets/xtvgie/face-datasets | 10779 |
| Shen et al. [16] (2017) | Automatic Acne Diagnosis | Binary classifier for skin detection and seven-class classifier for acne lesion classification, using CNNs with data augmentation techniques. | Binary classifier: 3000 cutaneous and 3000 non-cutaneous images. Seven-class classifier: 6000 images per class (blackheads, whiteheads, papules, pustules, cysts, nodules, normal skin). | Clinical images, augmented with rotation, shift, shear, scaling, and horizontal flip. | No | Not specified |
| Ayesha et al. [9] (2025) | Optimizing Acne Severity Detection | A transfer learning approach was implemented by extracting image features using a ResNet-152 pretrained model, then a fully connected layer was added and trained to learn the target severity level from labelled images. | DermNet NZ acne image sets used as the training dataset, and 100 test images were obtained to serve as a ‘golden set’ for model evaluation from Flicker Faces HQ Dataset . | The dataset acquired for the training process consisted of 1,148 selfie images of different people from different nationalities (European and South Asian), skin colors and features of mixed resolutions (high and low resolution) containing more than one acne lesion all over the face. | Datasets available at: https://www.kaggle.com/datasets/shubhamgoel27/dermnet | 1148 |
| Zhao et al. [17] (2019) | Acne Severity Assessment using Selfie Images | Deep learning model using transfer learning with ResNet 152 and | 4700 selfie images, with 4470 used for training and 230 used for testing. | Selfie images collected by Nestlé SHIELD and labeled by 11 dermatologists. | Code available on GitHub:https://github.com/ | 4700 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | a novel rolling image augmentation approach to improve CNN model generalization. | | | Microsoft/nestle-acne-assessment (accessed on 5 March 2025) | |
| Zhang et al. [13] (2024) | Advancements in acne detection | For the development of an advanced detection system, a multifunctional method for detecting acne using the CenterNet-based learning paradigm is proposed. | Through image usability testing and inclusion criteria, 153,183 acne lesions images were finally selected, from which a training set containing 150,219 images and a validation set containing 2,322 acne lesions images were constructed for subsequent modeling. | To train and validate our algorithm, we collected 212,374 original acne lesion images from dermatology clinical clinics in 15 hospitals in different provinces of China using smartphones from January 2020 to October 2022. | No | 212374 |
| Wu et al. [11] (2019) | Joint Acne Image Grading and Counting via Label Distribution Learning | Label Distribution Learning framework that integrates global grading and local lesion counting. | 1457 facial images from the ACNE04 dataset with 18,983 annotated lesions. | Clinical images collected under the Hayashi grading criterion, taken at a 70-degree angle. | Code and dataset available on GitHub: https://github.com/xpwu95/ldl (accessed on 5 March 2025) | 1457 |
| Alzahrani et al. [19] (2022) | Attention Mechanism Guided Deep Regression Model for Acne Severity Grading. | The developed architecture incorporates dilated UNet dense regressor for density regression with the information of bounding boxes generated from Faster R-CNN network, producing a hybrid detection–regression framework. | To conduct the experiments in this research work, a publicly available dataset named ACNE04 is used. | The ACNE04 contains 1457 images of lesions with 18,983 bounding boxes. | Code and dataset available on GitHub: https://github.com/xpwu95/LDL | 1457 |
| Seitè et al. [6] (2019) | Acne Grading from Smartphone Photographs | AI Algorithm trained on a large and diverse dataset, using deep learning techniques for acne grading and lesion identification. | 5972 images from 1072 patients with acne, including 2939 inflammatory lesions, 7603 non-inflammatory lesions, and 5702 PIHP instances. | Smartphone images (iOS and Android) across various racial groups. | No | 5972 |
| Phuc et al. [22] (2024) | ACNE8M - An acnes detection and differ- | Given the promising potential of YOLOv8, this model was selected for acne detection training. | Acnes detection dataset utilized for this model is sourced from Roboflow and authenticated by our dermatologists. | All images are resized to 800 x 800 pixels, with several augmentations applied: mosaic, horizontal and vertical flips, and | Datasets available at: | 9440 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | ential diagnosis system using AI technologies. | | | rotation augmentation within a range of -25 to 25 degrees. | 159 https://universe.roboflow.com/testdg2/fold_02 | |
| Lim et al. [18] (2019) | Automated Acne Grading using Deep Learning and CNNs | Deep learning models (DenseNet, Inception v4, ResNet18) trained on high-resolution images to automatically calculate IGA scores. | 472 frontal facial images from 416 acne patients, with training set of 314 images and testing set of 98 images, augmented to 6248 images. | Clinical images captured at ~ 2000 × 3000 pixel resolution. | No | 472 |
| Bernhard et al. [26] (2024) | Automatic Acne Severity Grading with a Small and Imbalanced Data Set of Low-Resolution Images. | A total of 1374 triplets of images (frontal and lateral views) from 391 different patients suffering from acne labeled with the IGA severity grade by an expert dermatologist were used to train and validate a deep learning model that predicts the IGA severity grade. | The study involved 391 Chinese patients | Pictures were taken with iPhone 11 with a ring light and a resolution of 3024×4032 pixels. | No | 1374 |
| Peris Fajarnés et al. [20] (2020) | Segmentation Methods for Acne Vulgaris Images using Fluorescence Imaging | K-means clustering algorithm implemented in MATLAB R2018b for automated segmentation and counting of acne lesions in fluorescence images. | 54 processed fluorescence images of mild acne patients, captured using iPhone X with Wood's lamp and LED lamp. | Fluorescence images captured using iPhone X smartphone. | No | 54 |
| Aldi et al. [28] (2025) | Multi-Class Facial Acne Classification using the EfficientNetV2-S Deep Learning Model. | Employed a quantitative experimental approach using facial acne image data, which were classified into five categories through a Convolutional Neural Network (CNN) model based on the EfficientNetV2-S architecture. | This dataset was obtained from a public repository on Kaggle, originally sourced from DermNet, and was preprocessed to support research in multi-class acne classification using digital facial images. | The dataset is divided into three subsets: 2,834 images for training, 921 for validation, and 918 for testing. | Datasets available at: https://www.kaggle.com/datasets/shubhamgoel27/dermnet/data | 4673 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Rashatapruck sa et al. [21] (2021) | Acne Detection | Faster R-CNN and R-FCN models used for acne detection, focusing on various types of acne lesions. | 871 annotated images with 15,917 lesions categorized into four types: Type I, Type III, PIE, and PIH. | Clinical images provided by Pan Rajdhevee Group Public Co., Ltd.; Bangkok, Thailand. | No | 871 |
| Yang et al. [23] (2021) | Acne Vulgaris Assessment | Inception-v3 architecture used for model construction, trained on clinical images categorized into four severity grades. | 5871 clinical images, with 1565 images in the training set, 392 in the validation set, and 40 in the test set. | Clinical images captured with single-lens reflex cameras. | No | 5871 |
| Watanabe et al. [31] (2025) | Deep Learning-Based Acne Severity Classification Using Standardized Facial Images of Japanese Patients. | An EfficientNet-B2 model was trained using a two-stage curriculum learning strategy, class weighting, and extensive data augmentation techniques. | A dataset of 349 facial images was collected from Japanese acne patients. | The patient cohort ranged in age from six to 76 years, with a median age of 21 and a mean age of 25.7 years. The dataset comprised 226 images from female patients (61.4%) and 142 from male patients (38.6%). | No | 349 |
| Quattrini et al. [15] (2022) | Facial Acne Classification | Used a VGG-19 based CNN for classification and BiSeNet for semantic segmentation, trained on re-annotated images from the FFHQ dataset. | 2307 re-annotated images from the Flickr-Faces-HQ (FFHQ) dataset, used for acne severity classification. | High-resolution images from the FFHQ dataset, re-annotated for acne severity. | Datasets available at: https://www.kaggle.com/datasets/arnaud58/flickrfaceshq-dataset-ffhq | 2307 |
| Min et al. [24] (2022) | Acne Detection using Mask-Aware Attention with Dynamic Context Enhancement | ACNet integrates Composite Feature Refinement, Dynamic Context Enhancement, and Mask-Aware Multi-Attention for enhanced feature representation and detection accuracy. | ACNE04 dataset with 1457 facial images and 18,983 annotated lesions. | Clinical photographs with varying resolutions, annotated by dermatologists. | Datasets available at: https://www.kaggle.com/datasets/manuelhettich/acne04 | 1457 |
| Mascarenhas et al. [36] (2025) | Improving acne severity detection. | Traditional methods to address blurring fail to recover fine details, making them unsuitable for teledermatology. To resolve this | This study utilized photographic images from four datasets: 1) the Acne Recognition Dataset 2) the CUHK dataset 3) the AR dataset 4) the XM2GTS dataset. | 228 people with skin problems rated 3-4 stage on the IGA scale took 3 selfies: frontal, two-thirds to the left, and two-thirds to the right. | Code and dataset available on https://github.com/Principiya1990/CATDeblurring | 650 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | issue, the study proposes a framework based on generative networks. | | | | |
| Zhang et al. [25] (2022) | Acne Detection | Ensemble neural network composed of ResNet50-based classification and YOLOv5-based localization modules for simultaneous severity, count, and bounding-box prediction | ACNE04: 1457 facial images with expert-annotated lesion counts, severity grades, and bounding boxes | ACNE04 dataset | Datasets available at: https://www.kaggle.com/datasets/manuelhettich/acne04 | |
| Wang et al. [29] (2022) | Cell Phone App for Facial Acne Severity Assessment | Acne-RegNet, a lightweight model for lesion detection and classification, combined with metadata for personalized severity assessment. | 1455 images from ACNE04 and 1515 images from a private dataset at Xiangya Hospital, processed into $80 \times 80$ lesion patches and resized for the model. | Clinical photographs, including metadata, collected from a private dataset and the ACNE04 dataset. | Datasets available at: https://www.kaggle.com/datasets/manuelhettich/acne04 | 2970 |
| Islam et al. [40] (2023) | Acne Vulgaris Detection and Classification | First, two CNN-based units are designed to extract deep feature maps, later combined in a feature aggregation module. The aggregated features provide rich input information and classify the acne by a softmax. | The authors have captured seventy-seven images of this dataset from the subjects who visited the Department of Dermatology at Bangabandhu Sheikh Mujib Medical University (BSMMU) and Dhaka Medical College (DMC). | The informed consent was obtained from all subjects before capturing the acne images by the 13-MP smartphone camera. | Datasets available at http://www.derm.net https://dermnetnz. org | 420 |
| Wang et al. [27] (2023) | Acne Detection and Severity Quantification | Two-stage deep learning scheme | 1803 facial images (1374 for training/test, 429 for clinical validation) | Smartphone images captured with iPhone 5s at Xiangya Hospital, annotated by dermatologists | No | 1803 |
| Huynh et al. [30] (2022) | Automatic Acne Object Detection and Severity Grading Using | Faster R-CNN deep learning model for acne object detection and LightGBM machine learning | 1572 facial images with 41,859 labeled lesions, captured by iOS and Android smartphones. | Smartphone images collected via the 'Skin Detective' mobile application. | No | 1572 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Smartphone Images and AI | model for grading acne severity based on the IGA scale. | | | | |
| Lin et al. [32] (2022) | Unified Acne Grading Framework on Face Images | KIEGLFN framework with Key Information Enhancement and Global Local Fusion Network, including a transfer fine-tuning strategy for new grading criteria. | 1457 images from the ACNE04 dataset, 200 images from the PLSBRACNE01 dataset, annotated for lesion location, severity, and type. | Clinical images from the ACNE04 and PLSBRACNE01 datasets. | Code available on GitHub: https:// github.com/linyi0604/KIEGLFN (accessed on 5 March 2025) | 1657 |
| Liu et al. [33] (2022) | Ensemble Pruning of Deep Learning Base Models for Acne Grading | Ensemble pruning of deep learning models, fivefold cross-validation, and transfer learning applied to both acne and skin cancer datasets. | 1457 images from the ACNE04 dataset, annotated for lesion location and severity grading; 3297 dermoscopic images in the skin cancer dataset, categorized as benign or malignant. | Clinical images from the ACNE04 and skin cancer datasets. | Datasets available at: https://github.com/xpwu95/ldl and https://www.kaggle. com/datasets/ fanconic/skin-cancer-malignant-vs-benign (accessed on 5 March 2025) | 4754 |
| Wen et al. [34] (2022) | Acne Detection and Severity Evaluation | Interpretable CNN-based object detection models, with a focus on interpretability | 1457 images from the ACNE04 dataset, with 1222 high-quality images used after data cleaning. | Clinical images with annotations for lesion location, severity, and bounding boxes. | Code available on GitHub: https:// github.com/wenh06/acne_detection/ and https://github.com/wenh06/yolov4_acne_torch | 1457 |
| Kim et al. [35] (2023) | Acne Lesion Detection and Counting for Severity Evaluation | CNN-based algorithm, with a focus on binary and five-class classification of acne lesions. | 1213 images with 20,699 labeled lesions (closed comedones, open comedones, papules, nodules/cysts, pustules). | Clinical images acquired in standardized settings using Canon EOS 550D and NIKON D7100 cameras. | Detailed code and hyperparameters provided in the Electronic Supplementary Materials. | 1213 |
| Wei et al. [37] (2023) | Accurate Acne Detection via Decoupled Sequential Detection Head | Decoupled Sequential Detection Head mechanism applied to mainstream two-stage detectors. | ACNE-DET: 276 high-resolution images with 31,777 labeled lesions; ACNE04: 1457 images with 18,983 bounding box annotations. | Clinical images acquired with VISIA system and annotated by dermatologists. | Code and dataset are publicly available, repository not specified. | 276 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Chen et al. [44] (2025) | Feature Feedback-Based Pseudo-Label Learning for Multi-Standards in Clinical Acne Grading. | Feature Feedback-Based Pseudo-Label Learning (FF-PLL) (1) an acne feature feedback (AFF) (2) all-facial skin segmentation (AFSS) (3) the AcneAugment (AA) strategy | This study proposes two different standard acne grading datasets, namely ACNE04 and ACNE-ECKH, to validate the effectiveness of the proposed method. | ACNE-ECKH dataset comprises 8242 images for acne image classification, consisting of 1500 labeled images and 6742 | Datasets available at: https://www.kaggle.com/datasets/manuelhettich/acne04 | 9699 |
| Li et al. [38] (2023) | AI-Powered Acne Grading System with Lesion Identification | ResNet50 architecture, with 945 images used for training and 240 for testing, incorporating lesion identification and severity grading. | 1501 standardized photos from 1501 acne patients, with detailed labeling of 15,922 lesions across 10 categories. | Clinical images obtained using the VISIA complexion analysis system. | No | 1501 |
| Lin et al. [39] (2023) | Acne Severity Grading via Diagnostic Evidence Distillation | Teacher-student structure using CNNs for both global estimation and lesion detection, integrating lesion counting and grading. | ACNE04: 1457 facial images with 13,633 lesions; PLSBRACNE01: 200 images categorized by severity. | Clinical images labeled according to Hayashi and Pillsbury criteria. | Code available on GitHub:https://github.com/linyi0604/DED (accessed on 5 March 2025) | 1457 |
| Li et al. [7] (2023) | AI Analysis of Dark Circles and Facial Skin Aging in Chinese Population | CNNs based system for grading facial features from selfies, validated against dermatologist assessments. | Over 1.9 million selfies from 1,939,586 participants, graded for facial aging indicators. | Selfies taken with smartphones, collected via a mobile app. | No | 1,939,586 |
| Li et al. [41] (2024) | Acne severity assessment | CNN-based automatic grading | 1,156,703 women aged 18-85 | Selfies taken with smartphones, collected via a mobile app. Annotated by dermatologists. | No | 1,156,703 |
| Kim et al. [42] (2024) | Facial Acne Segmentation | Semi-supervised learning using bidirectional copy-paste and deep learning with U-Net. | Acne images cropped from facial images | Clinical images (likely taken with specialized skin diagnostic equipment) | No | 2000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Zhang et al. [43] (2024) | Acne Detection using CenterNet | CenterNet-based deep learning framework, utilizing deep hierarchical aggregation strategies for feature extraction, combined with anchor-free detection for improved accuracy in locating and identifying acne lesions. | 153,183 images collected from 15 hospitals in different provinces of China using smartphones. | A mix of clinical and smartphone images | No | 153,183 |
| Prokhorov et al. [45] (2024) | Acne Image Grading using Label Distribution Smoothing | Gaussian label distribution generation and label smoothing techniques combined with a scale-adaptive approach, validated on the ACNE04 dataset using ResNet-50 architecture. | ACNE04 dataset with 1457 images and 18,983 bounding boxes of lesions. | Clinical images from the ACNE04 dataset, specifically annotated for acne grading. | Source code is publicly available at http://github.com/ open-face-io/acnelds | 1457 |
| Zein et al. [14] (2024) | Acne Dataset Generation Using Generative Adversarial Networks (GANs) | GAN-based method with StyleGAN2 to generate synthetic acne images at different severity levels, used for training a CNN-based classification system. | 1473 real images initially collected, with additional synthetic images generated across Mild, Moderate, and Severe severity levels. | Synthetic images generated using GANs based on clinical images. | Code and dataset available on GitHub: https: //github.com/hz0 1/AcneGAN/tree/ main | 1473 |
| Gao et al. [46] (2025) | Acne lesion detection and severity grading in online and offline scenarios | Deep learning model (AcneDGNet) that combines a vision transformer-based feature extractor with a CNN-based lesion detection module and a severity grading module | 2157 facial acne images compiled from two public datasets (ACNE04, AcneSCU) and three self-built datasets (AcnePA1, AcnePA2, AcnePKUIH) | Mixed capture sources: digital cameras, VISIA system, and smartphone images | Datasets available at: https://www.kaggle.com/datasets/manuelhettich/acne04 | 2157 |

## 4 Conclusions

Artificial intelligence-based acne assessment systems are expected to improve dermatological care by providing standardized assessments, reducing variability among observers, and improving clinical effectiveness. However, data discrepancies, limited data set diversity, and difficulties in integrating these tools into existing clinical work processes continue to limit their widespread use.

The widespread use of the method for determining the degree of acne based on artificial intelligence can standardize dermatological studies, allow for further comparison of studies conducted in different regions and clinical conditions, and facilitate the development of treatment methods based on reliable and repeatable data. In addition, the ability of artificial intelligence to process large amounts of data from different skin types can expand our understanding of the spread and geographical distribution of acne. However, although some models have demonstrated high accuracy under controlled conditions [46], the implementation of these results into clinical practice remains a complex task: to date, none of the proposed algorithms have been clinically confirmed, and no research on real-life situations has been published. To confirm the effectiveness of these artificial intelligence systems in everyday dermatological practice, it is necessary to conduct prospective studies with testing in real conditions.

Unlike other dermatological diseases, which have a broad public database supporting the development of a standardized artificial intelligence algorithm, acne is rare in these sources [47]. Confidentiality issues associated with the mass sharing of images of faces affected by acne remain a major obstacle to the creation of large, open datasets. Consequently, most datasets used in studies related to acne are not publicly available [15]. Quattrini et al. [15] used a public data set of faces: Flickr-Faces-HQ (FFHQ), consisting of 70,000 high-quality PNG images of human faces measuring 1024×1024 and containing significant differences in terms of age, ethnicity, and image background. However, since most images did not show acne, the data set was very disproportionate: approximately 88% of images fell under the "No acne" category. This discrepancy limited the study to determining the presence of acne rather than assessing its severity.

In conclusion, none of the published algorithms to date demonstrated the generality, reproducibility, or external and prospective clinical validation necessary for integration into conventional dermatological practice. Future work on acne diagnostics and severity assessment based on artificial intelligence should address these gaps, in particular, the need for prospective real-world clinical trials to assess model effectiveness in real-world settings[46]; the creation of large, publicly accessible and demographically diverse imaging repositories[14]; the study of federated education, allowing for the training of interagency models to maintain confidentiality; and the introduction of intelligible artificial intelligence methods to enhance clinicians' ability to interpret and perceive. Despite current limitations, the results are encouraging: AI-based tools promise to fundamentally change acne research and clinical care, reduce variability between and within observers, and provide truly personalized, timely treatment. By improving

model architectures, expanding and diversifying the existing data set, and firmly confirming effectiveness in real conditions, the industry can reveal the full potential of artificial intelligence in these conditions.

## References


1. Chen, H., Zhang, T.C., Yin, X.L., Man, J.Y., Yang, X.R., Lu, M.: Magnitude and temporal trend of acne vulgaris burden in 204 countries and territories from 1990 to 2019: An analysis from the Global Burden of Disease Study 2019. *Br. J. Dermatol.* 186, 673–683 (2021).
2. Reynolds, R.V., Yeung, H., Cheng, C.E., Cook-Bolden, F., Desai, S.R., Druby, K.M., Freeman, E.E., Keri, J.E., Gold, L.F.S., Tan, J.K., et al.: Guidelines of care for the management of acne vulgaris. *J. Am. Acad. Dermatol.* 90, 1006.e1–1006.e30 (2024).
3. Sharmin, S., Farid, F.A., Jihad, M., Rahman, S., Uddin, J., Rafi, R.K., Hossan, R., Karim, H.A.: A Hybrid CNN Framework DLI-Net for Acne Detection with XAI. *J. Imaging* 11, 115 (2025).
4. Bae, I.H., Kwak, J.H., Na, C.H., Kim, M.S., Shin, B.S., Choi, H.: A Comprehensive Review of the Acne Grading Scale in 2023. *Ann. Dermatol.* 36, 65–73 (2024).
5. Nichols, J.A., Chan, H.W.H., Baker, M.A.B.: Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophys. Rev.* 11, 111–118 (2018).
6. Seité, S., Khammari, A., Benzaquen, M., Moyal, D., Dréno, B.: Development and accuracy of an artificial intelligence algorithm for acne grading from smartphone photographs. *Exp. Dermatol.* 28, 1252–1257 (2019).
7. Li, J., Du, D., Zhang, J., Liu, W., Wang, J., Wei, X., Xue, L., Li, X., Diao, P., Zhang, L., et al.: Development and validation of an artificial intelligence-powered acne grading system incorporating lesion identification. *Front. Med.* 10, 1255704 (2023).
8. Jeong, H.K., Park, C., Henao, R., Kheterpal, M.: Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations. *JID Innov.* 3, 100150 (2022).
9. Singhapathiranage, A.P., Kelaart, D.D.K., Thathsarani, A.A.M., Liyanage, C.D.L., Lankasena, B.N.S.: Optimizing Acne Severity Detection: A Deep Learning Approach with Electronic Medical Record System Integration and Diverse Image Data. *Adv. Technol.* 5(1) (2025).
10. Van den Akker, O.R., Peters, G.Y., Bakker, C., Carlsson, R., Coles, N.A., Corker, K.S., Feldman, G., Moreau, D., Nordström, T., Pickering, J.S., et al.: Generalized Systematic Review Registration Form. Available at: https://osf.io/e8rhd (last accessed March 2025).
11. Wu, X., Wen, N., Liang, J., Lai, Y.-K., She, D., Cheng, M.-M., Yang, J.: Joint Acne Image Grading and Counting via Label Distribution Learning. In: *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Seoul, Republic of Korea, pp. 10641–10650 (2019).
12. Zhang, J., Zhang, L., Wang, J., Wei, X., Li, J., Jiang, X., Du, D.: SA-RPN: A Spatial Aware Region Proposal Network for Acne Detection. *IEEE J. Biomed. Health Inform.* 27, 5439–5448 (2023).
13. Zhang, D., Li, H., Shi, J., Shen, Y., Zhu, L., Chen, N., Wei, Z., Lv, J., Chen, Y., Hao, F.: Advancements in acne detection: application of the CenterNet network in smart dermatology. *Front. Med.* 11, 1344314 (2024).
14. Zein, H., Chantaf, S., Fournier, R., Nait-Ali, A.: Generative adversarial networks for anonymous acneic face dataset generation. *PLoS ONE* 19, e0297958 (2024).
15. Quattrini, A., Boër, C., Leidi, T., Paydar, R.: A Deep Learning-Based Facial Acne Classification System. *Clin. Cosmet. Investig. Dermatol.* 15, 851–857 (2022).

16.Shen, X., Zhang, J., Yan, C., Zhou, H.: An Automatic Diagnosis Method of Facial Acne Vulgaris Based on Convolutional Neural Network. *Sci. Rep.* 8, 5839 (2018).
17.Zhao, T., Zhang, H., Spoelstra, J.: A Computer Vision Application for Assessing Facial Acne Severity from Selfie Images. *arXiv preprint* arXiv:1907.07901 (2019).
18.Lim, Z.V., Akram, F., Ngo, C.P., Winarto, A.A., Lee, W.Q., Liang, K., Oon, H.H., Thng, S.T.G., Lee, H.K.: Automated grading of acne vulgaris by deep learning with convolutional neural networks. *Skin Res. Technol.* 26, 187–192 (2019).
19.Alzahrani, S., Al-Bander, B., Al-Nuaimy, W.: Attention Mechanism Guided Deep Regression Model for Acne Severity Grading. *Computers* 11, 31 (2022).
20.Fajarnés, G.P., Santonja, M.M., García, B.D., Lengua, I.L.: Segmentation methods for acne vulgaris images: Proposal of a new methodology applied to fluorescence images. *Skin Res. Technol.* 26, 734–739 (2020).
21.Rashataprucksa, K., Chuangchaichatchavarn, C., Triukose, S., Nitinawarat, S., Pongpruttthipan, M., Piromsopa, K.: Acne Detection with Deep Neural Networks. In: *Proc. 2nd Int. Conf. Image Processing and Machine Vision*, Bangkok, Thailand (2020).
22.Nguyen, P.K., Le, T.D., Nguyen, B.A., Nguyen, P.A.: ACNE8M – An acne detection and differential diagnosis system using AI technologies. *VNUHCM J. Sci. Technol. Dev.* 27(3), 3550–3561 (2024).
23.Yang, Y., Guo, L., Wu, Q., Zhang, M., Zeng, R., Ding, H., Zheng, H., Xie, J., Li, Y., Ge, Y., et al.: Construction and evaluation of a deep learning model for assessing acne vulgaris using clinical images. *Dermatol. Ther.* 11, 1239–1248 (2021).
24.Min, S., Kong, H., Yoon, C., Kim, H.C., Suh, D.H.: Development and evaluation of an automatic acne lesion detection program using digital image processing. *Skin Res. Technol.* 19, e423–e432 (2012).
25.Zhang, H., Ma, T.: Acne Detection by Ensemble Neural Networks. *Sensors* 22, 6828 (2022).
26.Bernhard, R., Bletterer, A., Le Caro, M., et al.: Automatic Acne Severity Grading with a Small and Imbalanced Data Set of Low-Resolution Images. *Dermatol. Ther.* 14, 2953–2969 (2024).
27.Wang, J., Wang, C., Wang, Z., Hounye, A.H., Li, Z., Kong, M., Hou, M., Zhang, J., Qi, M.: A novel automatic acne detection and severity quantification scheme using deep learning. *Biomed. Signal Process. Control* 84, 104803 (2023).
28.Aldi, Y., Kusnawi.: Multi-Class Facial Acne Classification using the EfficientNetV2-S Deep Learning Model. *AVITEC* 7, 299–309 (2025).
29.Wang, J., Luo, Y., Wang, Z., Hounye, A.H., Cao, C., Hou, M., Zhang, J.: A cell phone app for facial acne severity assessment. *Appl. Intell.* 53, 7614–7633 (2022).
30.Huynh, Q.T., Nguyen, P.H., Le, H.X., Ngo, L.T., Trinh, N.-T., Tran, M.T.-T., Nguyen, H.T., Vu, N.T., Nguyen, A.T., Suda, K., et al.: Automatic Acne Object Detection and Acne Severity Grading Using Smartphone Images and Artificial Intelligence. *Diagnostics* 12, 1879 (2022).
31.Watanabe, K., Iinuma, K., Nakashima, C., et al.: Deep Learning-Based Acne Severity Classification Using Standardized Facial Images of Japanese Patients. *Cureus* 17(9), e91944 (2025).
32.Lin, Y., Jiang, J., Ma, Z., Chen, D., Guan, Y., You, H., Cheng, X., Liu, B., Luo, G.: KIEGLFN: A unified acne grading framework on face images. *Comput. Methods Programs Biomed.* 221, 106911 (2022).
33.Liu, S., Fan, Y., Duan, M., Wang, Y., Su, G., Ren, Y., Huang, L., Zhou, F.: AcneGrader: An ensemble pruning of the deep learning base models to grade acne. *Skin Res. Technol.* 28, 677–688 (2022).

34.Wen, H., Yu, W., Wu, Y., Zhao, J., Liu, X., Kuang, Z., Fan, R.: Acne detection and severity evaluation with interpretable convolutional neural network models. *Technol. Health Care* 30, 143–153 (2022).
35.Kim, D.H., Sun, S., Cho, S.I., Kong, H.-J., Lee, J.W., Lee, J.H., Suh, D.H.: Automated Facial Acne Lesion Detecting and Counting Algorithm for Acne Severity Evaluation and Its Utility in Assisting Dermatologists. *Am. J. Clin. Dermatol.* 24, 649–659 (2023).
36.Mascarenhas, P.P., Sannidhan, M.S., Pinto, A.J., Camero, D., Martis, J.E.: Improving acne severity detection: a GAN framework with contour accentuation for image deblurring. *Front. Bioinform.* 5, 1485797 (2025).
37.Wei, X., Zhang, L., Zhang, J., Wang, J., Liu, W., Li, J., Jiang, X.: Decoupled Sequential Detection Head for accurate acne detection. *Knowledge-Based Syst.* 284, 111305 (2023).
38.Li, T., Ma, X., Li, Z., Yu, N., Song, J., Ma, Z., Ying, H., Zhou, B., Huang, J., Wu, L., et al.: Artificial intelligence analysis of over a million Chinese men and women reveals level of dark circle in the facial skin aging process. *Skin Res. Technol.* 29, e13492 (2023).
39.Lin, Y., Jiang, J., Chen, D., Ma, Z., Guan, Y., Liu, X., You, H., Yang, J.: DED: Diagnostic Evidence Distillation for acne severity grading on face images. *Expert Syst. Appl.* 228, 120312 (2023).
40.Islam, M.B., Junayed, M.S., Sadeghzadeh, A., Anjum, N., Ahsan, A., Shah, A.F.M.S.: Acne Vulgaris Detection and Classification: A Dual Integrated Deep CNN Model. *Informatica* 47(4) (2023).
41.Li, T., Ma, X., Li, Z., Yu, N., Song, J., Ma, Z., Ying, H., Zhou, B., Huang, J., Wu, L., et al.: Facial adult female acne in China: An analysis based on artificial intelligence over one million. *Skin Res. Technol.* 30, e13693 (2024).
42.Kim, S., Yoon, H., Lee, J.: Semi-Supervised Facial Acne Segmentation Using Bidirectional Copy–Paste. *Diagnostics* 14, 1040 (2024).
43.Zhang, J., Zhang, L., Wang, J., Wei, X., Li, J., Jiang, X., Du, D.: Learning High-quality Proposals for Acne Detection. *arXiv preprint* arXiv:2207.03674 (2022).
44.Chen, Y.-Y., Chan, H.-T., Wang, H.-C., Wang, C.-S., Chen, H.-H., Chen, P.-H., Chen, Y.-J., Hsu, S.-H., Hsia, C.-H.: Feature Feedback-Based Pseudo-Label Learning for Multi-Standards in Clinical Acne Grading. *Bioengineering* 12, 342 (2025).
45.Prokhorov, K., Kalinin, A.A.: Improving Acne Image Grading with Label Distribution Smoothing. In: *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI)*, Athens, Greece, pp. 1–5 (2024).
46.Gao, N., Wang, J., Zhao, Z., Chu, X., Lv, B., Han, G., Ni, Y., Xie, G.: Evaluation of an acne lesion detection and severity grading model for Chinese population in online and offline healthcare scenarios. *Sci. Rep.* 15, 1119 (2025).
47.Choy, S.P., Kim, B.J., Paolino, A., Tan, W.R., Lim, S.M.L., Seo, J., Tan, S.P., Francis, L., Tsakok, T., Simpson, M., et al.: Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. *npj Digit. Med.* 6, 180 (2023).
48.Haggenmüller, S., Schmitt, M., Krieghoff-Henning, E., Hekler, A., Maron, R.C., Wies, C., Utikal, J.S., Meier, F., Hobelsberger, S., Gellrich, F.F., et al.: Federated Learning for Decentralized Artificial Intelligence in Melanoma Diagnostics. *JAMA Dermatol.* 160, 303–311 (2024).

# From Caspian Shores to Mountain Peaks: ICT-Powered Tourism Resilience in Azerbaijan

Shahla Gahramanova[1][0000-0002-8317-3913] and Kamal Ibrahimov[2][0009-0005-7376-5381]

[1] Azerbaijan University of Architecture and Construction, Baku, A. Sultanova,11, Azerbaijan
shahlakahramanova@gmail.com
[2] Azerbaijan University of Architecture and Construction, Baku, A. Sultanova,11, Azerbaijan
k.ibrahimov@gmail.com

**Abstract.** Azerbaijan's diverse geography – from mountain resorts and forested national parks to coastal areas along the Caspian Sea and historic urban centers – makes the country increasingly vulnerable to climate-related risks that affect tourism. The country spans 9 out of 11 major climate types, which generates highly varied tourism patterns, including more than 2 million annual visitors to Baku and approximately 250-300 thousand to major mountain resorts such as Shahdag and Gabala. These hazards – including heatwaves, landslides, flooding, coastal storms and pressures on heritage sites pose challenges to visitor safety and long-term sustainable development. This paper examines how information and communication technologies (ICT) can strengthen tourism risk management and climate adaptation in Azerbaijan, including regions with still-developing digital infrastructure. The paper highlights a range of accessible ICT tools (weather monitoring platforms, mobile alert systems, open-source GIS mapping and digital route guidance) that can support early warning, visitor management and spatial planning. Special attention is given to the role of crowdsourced information, including tourist feedback, geotagged images and community-based reporting. These sources can provide valuable insights even in the absence of costly technological systems. Based on this analysis, the paper proposes a practical model for integrating ICT into local governance and destination management in Azerbaijan. The findings suggest that even simple digital solutions can enhance climate resilience, protect cultural and natural heritage and improve the overall safety and experience of visitors. This approach supports Azerbaijan's broader strategy to develop smart, sustainable and competitive tourism destinations.

**Keywords:** Azerbaijan Tourism, Climate Adaptation, Risk Management, ICT Tools, Resilient Tourism Planning, Sustainable Destinations

## 1 Introduction

Tourism is one of the fastest-growing sectors of the global economy, significantly contributing to employment, income generation and cultural exchange. However, this industry is highly sensitive to climate change, as it is dependent on natural landscapes, seasonal weather patterns and environmental quality [1]. Extreme heat, floods and un-

predictable weather events can disrupt tourism activities, reduce visitor safety and negatively impact the socioeconomic stability of host territories. Research shows that tourism is often more vulnerable to climate change than many other economic sectors, highlighting the need to develop and implement proactive adaptation strategies [1].

In this context, the example of Azerbaijan, a country with exceptional natural and climatic diversity and a wide range of tourism resources, is particularly illustrative. Situated at the crossroads of Europe and Asia, Azerbaijan encompasses nine of the eleven main climate types, ranging from the subtropical coastal plains along the Caspian Sea to the snow-capped highlands of the Greater Caucasus. This diversity provides the basis for the development of various tourism segments: ski and mountain resorts, beach holidays, ecotourism and cultural and historical routes. In particular, the mountain resorts of Shahdag and Gabala, beach destinations on the Absheron Peninsula, ecotourism in the forests of Lankaran and cultural centers such as Icheri Sheher, Shusha and Sheki generate a stable tourist flow [2].

A well-developed system of specially protected areas adds additional value to the country's tourism potential. Key sites include the Shirvan National Park, Absheron National Park, Hirkan National Park and the Gizilagach biosphere reserve. Collectively, these areas occupy more than 10% of the country's territory and include internationally recognized ecosystems [3]. As tourism activity near these areas increases, the need to integrate climate change adaptation and environmentally sound management principles into spatial planning processes grows.

Climate challenges in Azerbaijan are exacerbated by natural, geographical and geodynamic factors. Approximately 25-30% of the country's territory is prone to landslides, particularly on the slopes of the Greater and Lesser Caucasus and adjacent foothills, posing risks to the development of mountain tourism clusters and infrastructure. Additionally, approximately 1300 $km^2$ of territory is in a potential flood zone due to river flooding and fluctuating water levels. Furthermore, the entire country is located within a seismically active zone [4; 5]. Thus, tourism development in Azerbaijan is taking place in the context of complex climate and geohazard sensitivity, requiring a systemic and interdisciplinary approach to spatial planning.

The geographic specificity of tourist destinations exacerbates the significance of these risks. Mountain resorts such as Shahdag (1300-2435 m) and Gabala (600-1500 m) face challenges with unstable snow cover and landslide activity. Meanwhile, the coastal areas of the Absheron Peninsula, with a coastline of approximately 800 km are experiencing increasing exposure to heat waves, storm surges and coastal erosion. Tourist flows reflect this spatial diversity: Baku hosts approximately 2 million foreign tourists annually, Shahdag attracts approximately 150-200 thousand seasonal visitors and Lankaran attracts 150-200 thousand tourists in the summer season (table 1.). The concentration of tourist flows in vulnerable areas increases the pressure on infrastructure and highlights the need to consider climate risks in regional planning.

Despite its wealth of natural and cultural resources. Azerbaijan's tourism sector faces increasing climate threats that could impact both the quality of the tourism product and its long-term sustainability. Rising average annual temperatures, changing precipitation patterns, more frequent extreme weather events and water shortages can negatively impact coastal resorts, mountain routes and cultural heritage sites. Scientific

research emphasizes that sustainable tourism development concepts should integrate climate factors into spatial and strategic planning systems, combining environmental and socio-economic aspects [6].

**Table 1.** Major tourism clusters in Azerbaijan and their climate-risk profile.

| Tourism cluster | Core tourism activities | Key climate and geohazards | Visitor load (annual) | Planning priorities |
|---|---|---|---|---|
| Shahdag (Gusar) | Winter sports, mountain tourism | Snow variability, avalanches, landslides | 250-300 thds. | Slope stabilization, climate adaptive ski infrastructure |
| Gabala | Resorts, adventure parks, nature tourism | Landslides, heavy precipitation, forest fire risk | 200-250 thds. | Hazard-sensitive zoning, forest-interface management |
| Sheki-Zagatala | Heritage tourism, mountain villages | Landslides, flash floods | 120 thds. | Protection of heritage zones, flood-risk spatial planning |
| Khinaliq (Guba) | Highland ecorouts, rural tourism | Avalanches, road closures, erosion | 80-100 thds. | Climate-proof road network, controlled visitor access |
| Baku (Absheron) | Urban tourism, cultural heritage, coastal leisure | Heatwaves, coastal erosion, flooding | 2 mln. | Coastal protection, heat-stress mitigation, digital visitor management |
| Lankaran, Lerik | National parks, humid forest | Flooding, landslides | 150-200 thds. | Buffer zones around protected areas, slope monitoring |
| Naftalan | Heath tourism | Heatwaves | 50-70 thds. | Climate-resilient health facilities |
| Goygol | Lake and national park tourism | Landslide, erosion | 60-80 thds. | Visitor-capacity control, ecosystem protection |

In response to these challenges, information and communication technologies (ICT) are playing an increasingly significant role as a tool for increasing the adaptability of tourist areas. Even without the implementation of a complex technological infrastructure – such as large-scale IoT sensor networks or centralized climate databases – ICT-based solutions can significantly support risk management processes. These include weather monitoring systems, open geographic information platforms, mobile alert systems and crowdsourced data collection mechanisms [7].

Given the high level of digital accessibility in the country – over 80% mobile penetration and approximately 79% Internet coverage – Azerbaijan has the necessary foundation for implementing such tools. This creates the preconditions for engaging both large tourist destinations and small operators in basic monitoring processes, participation in GIS risk assessments and the use of mobility warning systems.

Thus, in the face of growing climatic and geo-ecological challenges, a conceptual rethinking of tourism management is needed. This study proposes a conceptual ICT-based risk management model adapted to the tourism landscape of Azerbaijan. It summarizes international practices and formulates strategies aimed at increasing the resilience of tourism destinations, protecting visitors and ensuring long-term sustainable development without relying on complex technological infrastructure.

## 2 Literature review

The integration of information and communication technologies (ICT) into the tourism industry has become a transformative force. ICT reshapes how destinations manage services, engage stakeholders and respond to emerging challenges. Numerous studies demonstrates that ICT supports data-driven decision-making and enhances strategic planning by improving operational efficiency and communication among industry actors [8]. These technologies influence all aspects of tourism management, from reservations and marketing to service delivery and visitor experience [9].

Within the broader context of sustainable development, ICT has also recognized as a key driver of environmental quality and resource efficiency. For instance, research in China indicates that long-term ICT adoption correlates with improved environmental indicators, as digital platforms and efficiency gains help mitigate pressures from increased tourism activity [7]. These findings support the global discourse digitalization's role in balancing tourism growth with ecological sustainability.

A growing body of scholarship emphasizes the integrating disaster risk reduction (DRR) with sustainable tourism. Systematic reviews show that DRR measures – from advance planning to early warning systems – are fundamental for building resilient tourism destinations and mitigating environmental impacts [10]. Embedding risk awareness into both policy and daily management protects visitors and local communities alike.

Although research on digital twins and advanced simulation technologies in tourism is still emerging, early studies indicate substantial potential. Digital frameworks can support destination management and help preserve cultural heritage at the site level [11]. As the field devolves, digital replicas of real-world tourist environments may offer new paths for planning and resilience modeling.

ICT can strengthen collective action and coordination among stakeholders responsible for sustainable tourism management. For instance, studies of remote Arctic tourism sites illustrate that digital tools help unify dispersed actors, improve communication and promote shared sustainable norms. However, researchers note that ICT effectiveness is influenced by users' technological capabilities and broader social capacity [12].

Despite notable progress in understanding ICT's role in tourism, significant gaps remain, particularly its application to climate-related risk management in destinations with limited technological infrastructure. Most studies focus on data-rich contexts, with fewer exploring how ICT can be adapted to emerging economies with nascent digital ecosystems. This underscores the need for contextually tailored models that leverage accessible ICT solutions to strengthen tourism resilience, especially in countries like

Azerbaijan, where tourism assets are vulnerable to climate variability and infrastructure is still developing.

Recent research in Azerbaijan is beginning to addressed this gap. Studies on regional tourism development illustrate how tourism growth can support local economic and community stability across rural and urban destinations [13]. Furthermore, research on the cultural tourism sector emphasizes the importance of heritage preservation, cultural products and diversified tourism offerings in enhancing competitiveness and promoting sustainable growth [14].

## 3 Methodology and conceptual framework

The proposed conceptual framework for ICT-supported tourism resilience in Azerbaijan builds on the principle that digital data, spatial analysis and analytics can improve destination planning, risk assessment and decision-making, even where advanced sensor networks are not yet widespread. International research shows that geographic information systems (GIS) and digital tools increasingly inform tourism planning, visitor behavior analysis and sustainable destination management. For example, GIS technologies have been successfully applied to map points of interest and analyze visitor flows, providing evidence to guide spatial planning in historic or environmentally sensitive areas [15].

At its core, the framework integrates spatial, behavioral and contextual data layers to develop a dynamic understanding of tourism flows and climate-related vulnerabilities. Spatial data include geographic and environmental information from open platforms such as OpenStreetMap, topographic models, land cover datasets and national GIS repositories. These layers help identify risk zones, accessibility patterns and the spatial distribution of tourism assets.

A second data category encompasses digital traces of tourism activity, such as aggregated mobile traffic patterns, web searches and social media presence. These data sources have proven valuable for analyzing tourism dynamics and forecasting demand [16]. Early research indicates that user-generated content, device data, transaction and booking information can reveal patterns in visitor behavior and support strategic planning. Even without extensive IoT infrastructure [16].

Additionally, studies on nature-based tourism highlight the role of mobile applications, geospatial platforms and interactive digital tools in managing visitor flows and supporting environmental sustainability [17]. These examples suggest that digital tools can extend beyond marketing and engagement into monitoring and planning and resilience-building functions.

From these foundations, the framework envisions integrated ICT applications, including GIS-based spatial analysis, digital dashboards, aggregated behavioral indicators and scenario modelling, can be aligned into a unified system for resilience planning. Accessibility is central: initial implementation relies on open or publicly available data, while future stages may incorporate real-time IoT sensor or advanced predictive models. Research on digital twin applications in tourism shows that these tools remain emergent and exploratory, not yet widely operational [11].

In practical terms, operationalizing this conceptual framework for Azerbaijan would begin with the development of an integrated GIS database of tourism assets, risk factors and climatic variability indicators. Simple analytical dashboards can visualize seasonal patterns and hazard probabilities. Aggregated behavioral data can highlight high-pressure zones and forecast tourist flows. This system allows local authorities, tourism agencies and community organizations to make data-informed decisions that enhance visitor safety, protect natural and cultural heritage and support sustainable destination development. As ICT capacity grows, additional layers, such as predictive analytics and machine learning models, can refine resilience strategies and align them with international best practices.

This approach combines spatial clustering of tourism regions, climate-risk profiling and an assesment of ICT applicability under varying infrastructural conditions. Such a composite method allows integrating environmental risk, territorial specificity and digital capacities into a single spatial-planning oriented analytical framework.

## 4 ICT Applications for climate risk management in tourism

The integration of digital technologies into tourism planning has gained substantial attention in recent years, driven by the increasing need to address climate-related risk and to strengthen destination resilience. ICT tools offer mechanisms for monitoring environmental conditions, informing visitors, analyzing mobility patterns and supporting early warning systems, all of which are critical for minimizing the impact of climate variability on tourism destinations.

Environmental monitoring systems represent of the most direct applications of ICT in climate risk management. These systems collect and interpret meteorological, hydrogeological and geospatial data to support early detection of hazards. International examples show that weather-monitoring platforms, remote sensing technologies and automated alert systems substantially improve preparedness and response capacity in tourism areas. For instance, studies demonstrate that digital early warning systems significantly reduce vulnerability to extreme weather events by enhancing real-time communication with tourists and local stakeholders [18].

In Azerbaijan, the State Agency for Hydrometeorology increasingly utilizes remote sensing products and short-term forecasting models warnings about floods, landslides and heatwaves – hazards that affect tourism hub such as Guba, Lankaran and Shahdag. While these systems are not yet fully integrated into tourism applications, they form a foundation for future ICT-enabled monitoring and warning tools tailored specifically for the tourism sector.

Spatial datasets generated through ICT tools can directly support urban planning decisions, such as defining buffer zones around heritage sites, mapping evacuation corridors, regulating tourist density in fragile landscapes or adapting building codes in hazard-prone territories. GIS-based tools allow destinations to visualize risk-exposed zones, evaluate accessibility routes and assess interactions between natural hazards and

tourism infrastructure. Research highlights the use of spatial risk models, climate exposure maps and multi-criteria decision-making for evaluating hazard vulnerability in tourism destinations [19].

Such GIS-based techniques are particularly relevant for Azerbaijan, where highly diverse geography and topography create differentiated spatial planning challenges. Mountain tourism clusters such as Gabala and Shahdag are exposed to landslide activity and snow-cover variability, while coastal destinations (including Absheron, Khachmaz and Lankaran) face sea-level fluctuation, shoreline erosion and intensifying heat stress. Historic urban centers – most notably Baku's Icheri Sheher, Sheki and Ordubad - require spatial planning approaches that accounts for microclimate impacts on architectural heritage and the cumulative pressures of visitor flows. International experience demonstrates that GIS-supported vulnerability mapping, widely applied in countries such as Italy, Japan and Norway, can be effectively adapted to these conditions to inform infrastructure decision and strengthen environmental management [15].

Understanding visitor mobility pattern is equally critical for anticipating exposure to climate-related hazards and managing congestion in vulnerable areas. ICT-enabled tools such as mobile positioning data, social media analytics and geotagged imagery have proven effective for examining the spatiotemporal dynamics of tourism. Existing research confirms that mobile big data can substantially enhance the accuracy of demand forecasting, congestion detection and flow prediction [20], while social media data - from platforms such as Instagram and Twitter -provide valuable information on visitor preferences, perceived risk and real-time responses to climate-induced disruptions [21]. These digital traces, when integrated into spatial decision-making frameworks, offer a scalable source of evidence for climate-sensitive planning in Azerbaijan's tourism regions.

For Azerbaijan, such approaches can offer valuable evidence into seasonal visitor flows in Naftalan, Shahdag ski resort and Goygol National park, heat-related behavioral changes in Baku's urban tourism areas and demand fluctuations in Lahich, Guba and Sheki during periods of unstable weather.

These data can support the development of adaptive scheduling, crowd management strategies and climate-informed marketing for tourism destinations.

Mobile applications have become essential for disseminating real-time information on hazard, weather alerts and safe routes. International studies show that location-based services enhance visitor safety by providing personalized alerts and route recommendations [22; 23; 24].

Globally, destinations such as New Zealand, Japan and Switzerland already use mobile apps for avalanches alerts, heatwave notifications, trail closures and flooding updates.

In Azerbaijan, where smartphone penetration exceeds 80%, developing lightweight apps or integrating tourism functions into existing state platforms (e.g., myGov, ASAN services, Meteo.az) could ensure cost-effective and scalable communication channels for climate-related risk information.

## 5 Discussion

The finding and conceptual propositions presented in this study highlight that ICT-supported resilience frameworks hold significant potential for strengthening tourism sustainability in Azerbaijan, but their implementation requires a nuanced understanding of both technological capacities and socio-institutional contexts. Consistent with global research, the analysis shows that geospatial data, digital communication tools and behavioral analytics can enhance the capacity of destinations to anticipate, respond to and recover from climate-induced risks [15]. The operationalization of these tools depends on factors including data availability, institutional readiness and disparities between urban and rural digital infrastructures. Given the spatial differentiation of Azerbaijan's tourism region – mountainous, coastal, forested and heritage-urban zones – the degree of climate exposure and planning requirements vary, reinforcing the need for an ICT-supported approach where digital monitoring, early-warning tools and GIS-based zoning align risk management with territorial planning priorities.

Approximately one-quarter of Azerbaijan's territory is landslide-prone, flood-exposed areas cover substantial river basins and the entire nation is seismically active. These hazards highlight the importance of spatially informed ICT integration into urban and regional planning [4; 5]. International studies show that ICT-driven solutions – particularly GIS-based risk maps, mobile early warning systems and big data analytics – are most effective when integrated into long-term destination management strategies [25]. Countries such as Japan, New Zealand and Italy demonstrate how multi-stakeholder digital platforms combine hazard monitoring, tourist mobility data and emergency communication [26]. Compared to Italy, where coastal tourist zones host millions of visitors annually, Azerbaijan's coastal destinations receive 250-300 thousand tourists each summer, indicating smaller scale but highly climate-sensitive management needs. ICT-based tools can support hazard-sensitive zoning, infrastructural adaptation and local-specific development strategies – elements that remain underdeveloped in Azerbaijan's planning practice.

At the same time, Azerbaijan possesses comparative advantages that can accelerate digital adoption, including high mobility phone penetration, expanding broadband coverage and e-government initiatives (e.g., ASAN, eGov portal) - providing a strong foundation for ICT applications in tourism resilience. Its diverse climatic structure, - nine climate zones - generates highly differentiated vulnerability patterns. Mountain areas such as Gabala, Gusar and Shahdag face snow-cover variability, landslides and glacial melt dynamics, where coastal areas including Absheron, Lankaran and Khachmaz are exposed to heat stress, shoreline erosion and episodic storm surges. Cultural heritage sites – most notably Sheki, Baku's Old city, Ganja and Ordubad – experience humidity variation, localized heat islands and rising visitor pressure, making digital monitoring platforms and predictive modeling valuable for conservation-oriented spatial planning.

By aligning ICT applications with these region-specific climatic patterns, Azerbaijan can formulate differentiated resilience strategies that correspond to the distinct environmental and cultural profiles of its tourism destinations. This approach also addresses

the limitation of generic climate adaptation frameworks commonly presented in international studies, demonstrating the need for localized, spatially sensitive and heritage-conscious digital solutions.

Compared to large-scale protected landscape in countries like Italy or New Zealand, Azerbaijan's preserved zones - from semidesert ecosystems of Shirvan National park to the humid forests of Hirkan National park - are similar in area but extremely diverse in ecological function. This increases the spatial planning challenge. So, even small disturbances like tourist inflow, erosion, unauthorized access, can create disproportionately large impact. Therefore ICT-supported climate risk tools should be integrated not only into tourism strategies, but also into land-use regulation around these preserved territories to ensure ecologically stability resilience.

A critical insight emerging from this study is that institutional collaboration is as important as technological capability. Tourism resilience requires coordination among the Ministry of Ecology and Natural Resources, the State Tourism Agency, local authorities and private tourism operations. Without shared data protocols and inter-agency platforms, ICT tools risk becoming fragmented.

Successful integration of ICT into climate-risk management requires embedding digital monitoring and geospatial analytics into regional master plans, general plans of tourism zones and protected-area management strategies.

Global evidence shows that destinations with strong governance mechanisms – such as those in Norway and South Korea – tend to adopt ICT-based risk tools more successfully. Azerbaijan's ongoing reforms in digital governance are promising, but tourism-specific integration remains limited. The absence of formal data-sharing mechanisms, limited training in GIS and analytics for regional tourism departments and insufficient involvement of local communities create obstacles that must be addressed before the full potential of digital resilience systems can be realized.

Another dimension revealed by the analysis related to equity and accessibility. While mobile phone penetration is high nationally, disparities persist between urban centers like Baku and remote mountain settlements. ICT-based systems may unintentionally reinforce inequalities if local tourism operators – guesthouses, ecolodges, small tour companies – lack the digital literacy necessary to participate in data sharing or use mobility dashboards.

From spatial planning perspective, unequal access to ICT between urban and rural territories may translate into uneven adaptive capacities, influencing regional development trajectories and resilience planning.

Research from developing regions highlights that ICT-driven resilience frameworks are effective only when accompanied by efforts to increase digital skills and ensure inclusive access. In Azerbaijan, this calls for targeted capacity-building programs, especially in Guba-Khachmaz, Sheki-Zagatala and Shirvan tourism clusters where small-scale tourism enterprises dominate.

Implementing the proposed framework can generate tangible benefits across institutional, economic and community domains. For policymakers, it enhances the capacity to delineate high-risk tourism zones, prioritize spatial investments and integrate climate-sensitive criteria into regional and urban planning processes. For tourism enterprises, the framework contributes to more reliable demand projections by incorporating

mobility and behavioral evidence, while also reducing operational vulnerabilities through earlier warnings and more flexible scheduling. For local communities, improved risk management strengthens the stability of tourism-dependent livelihoods by mitigating climate-related disruptions and provides new avenues for participation in monitoring and reporting mechanisms supported by ICT tools. These practical benefits reinforce that ICT-enabled resilience should be viewed not simply as a technological enhancement but as an essential strategic component of sustainable tourism development in Azerbaijan.

ICT-based resilience approaches are increasingly recognized as consistent with global sustainability agendas. Studies from China, Europe and the Pacific show that digitalization can promote environmentally responsible form of tourism by reducing resource inefficiencies, alleviating pressure in sensitive landscapes and improving the spatial coordination of land-use decisions [17]. In the context of Azerbaijan – where tourism development intersects with culturally significant sites and ecologically fragile territories – the use of ICT instruments can meaningfully contribute to broader national objectives, including the preservation of biodiversity within protected areas, the safeguarding of UNESCO-recognized heritage, the promotion of low-impact tourism practices and the reduction of extensive dependence on climate-sensitive seasonal activities. In this sense, digitalization functions not only as a mechanism for climate-risk mitigation but also as a catalyst for more sustainable and balanced territorial development [17].

Further research is nevertheless needed to strengthen the empirical basis for ICT-enabled resilience in Azerbaijan. Important directions include the application of the proposed framework to specific destinations such as Shahdag or Goygol, the integration of localized climate datasets from the Ministry of Ecology and Natural Resources with tourism indicators to refine spatial modelling and the exploration of digital twin methodologies for safeguarding cultural heritage sites in Baku and Sheki. Additional attention should be given to understanding community-level digital literacy, which may shape the feasibility of ICT uptake, as well as to evaluating the potential of mobile early-warning applications designed specifically for tourists. Advancing these research areas will support the development of more grounded, context-sensitive strategies for climate adaptation and digital transformation within Azerbaijan's tourism system.

## 6 Urban and spatial planning strategies for ICT-enhances tourism resilience

The findings of this study highlight that the integration of ICT into urban and regional planning is essential for building resilient tourism destinations in Azerbaijan. Spatial planning functions as a coordinating framework that links tourism management, climate adaptation and sustainable development, ensuring that growth is guided by coherent environmental, social and economic principles. Given the country's diverse geography – from the Caspian coastal plains to the high mountain systems of the Greater Caucasus – planning approaches must be sensitive to regional variability in climate exposure, topography and tourism dynamics.

Mountain destinations such as Gabala, Guba (including Shahdag and Khinalig) and Gusar are characterized by complex terrain and climate-sensitive conditions, including landscape-prone slopes, snow-cover fluctuations, and meltwater processes. In such settings, planners can rely on GIS-based slope analysis, hazard overlays and trail optimization models to inform the design of safe hiking routes, ski trails and transport access. For instance, the Shahdag ski resort could strengthen safety management by integrating snow-depth sensors with GIS-supported visitor-flow simulations, thereby reducing congestion and minimizing accident risks. Similarly, ecotourism areas in Gabala may benefit from seasonal runoff modelling and landslide susceptibility analysis, which can guide local zoning and infrastructure placement.

Coastal regions such as Absheron, Lankaran and Khachmaz face distinct pressures associated with shoreline erosion, heat stress and periodic storm surges. Spatial planning responses in these areas include coastal setback regulations for tourism infrastructure, buffer-zone designation and climate-informed strategies for urban expansion. Open-access GIS mapping of erosion-prone coastal segments on the Absheron peninsula can assist local authorities in development control, while real-time meteorological alerts and hazard maps embedded in tourist applications enhance visitor safety and preparedness.

Heritage-rich urban centers – including Baku's Icheri Sheher, Sheki, Ganja, Ordubad and the historical settlements of Ismayilli (Lahij and Basgal) – require planning approaches that balance conservation priorities with climate adaptation and visitor management. ICT-enabled monitoring systems can track visitor density, microclimatic variations and structural sensitivity of heritage buildings. For example, the Palace of the Sheki Khans would benefit from microclimate sensors integrated with GIS-based visitor tracing to mitigate pressure on fragile interiors, while urban planning could reinforce pedestrianization and adaptive routing strategies to reduce local congestion.

ICT tools can strengthen these spatial-planning strategies by improving the visibility, coordination and analytical capacity of decision-makers. GIS-based hazard and asset mapping supports real-time visualization of risk patterns and land-use constraints, while digital dashboards and scenario-modelling platforms enable simulations of tourism flows, urban growth trajectories and climate impacts [27]. Crowdsources data and mobile analytics deepen the understanding of visitor behavior and seasonal dynamics, informing adaptive zoning, mobility planning and operational readiness [28]. Integrating these digital resources into municipal and regional planning platforms ensures that tourism, environmental and urban datasets converge, reducing fragmentation and enhancing the coherence of planning decisions. These stages from a progressive, multi-level approach to ICT-enabled resilience that links spatial planning with digitalization (Table 2).

**Table 2.** ICT-integrated spatial planning model for climate-resilient tourism in Azerbaijan

| Adaptive stage | Planning focus | ICT-function | Typical tools/Application |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Reactive (Vulnerability is acute) | Short-term responses; hazard awareness; temporary restrictions in high-risk areas | Basic information flows; initial digital mapping; early notifications | SMS alerts, simple weather monitoring, isolated GIS risk maps, basic crowdsources reporting |
| Transitional (Vulnerability is chronic) | Medium-term adjustments; route optimization; early elements of climate-sensitive zoning | Integrated datasets; municipal tourism data exchange; monitoring of visitor flows | Mobile route apps with hazard information, open-source GIS in municipal workflows, crowdsourced hazard mapping, mobility pattern analytics |
| Transformative (Vulnerability is reduced) | Long-term strategic planning; resilient infrastructure design; climate-adaptive regional policies | Fully integrated digital ecosystem supporting predictive modelling and coordinated decision-making | Digital twins of tourism areas, AI-based risk forecasting, unified ICT planning platforms, real-time environmental sensors, data-driven zoning tools |

Institutional coordination remains crucial. Effective and sustainable ICT adoption requires close collaboration between urban planners, environmental agencies, tourism authorities and local communities. Shared spatial-data platforms can consolidate hazard information, mobility indicators and development constraints, preventing isolated decision-making and reinforcing resilience at multiple governance levels. Embedding climate adaptation criteria in zoning regulations, expanding the use of GIS in municipal workflows and promoting participatory planning where communities contribute to monitoring and data validation can further support integrated territorial management. Strategically differentiated resilience strategies – tailored separately for mountainous areas, coastal zones and heritage cities – are essential to balance visitor management with environmental and cultural preservation.

Advantages arising from this integrated approach are evident across different territorial categories. In mountain regions, ICT-supported planning can reduce landslide exposure, optimize ski and ecotourism infrastructure and enhance safety. In coastal zones, climate-informed land-use strategies strengthen protection against erosion and flooding while improving heat mitigation. In heritage cities, digital monitoring contributes to reducing visitor pressure, adapting public spaces to climatic stressors and safeguarding culturally significant buildings.

For the preserved territories of Azerbaijan, spatial planning should prioritize climate-buffer corridors, impact-controlled visitation routes and strict construction limitations near sensitive ecosystems. Ecosystem-oriented development guidelines are particularly relevant for highland parks, where erosion risks are elevated. Embedding ICT-based environmental and visitor monitoring into these planning processes ensures that tourism development remains compatible with biodiversity protection and long-term climate resilience.

## 7 Conclusion

This study examines how ICT-supported frameworks can enhance climate resilience within Azerbaijan's tourism sector, emphasizing the intersections between tourism planning, urban development and spatial management. The results demonstrate that even in the absence of extensive IoT networks or full-scale digital twin systems, Azerbaijan can effectively rely on accessible geospatial tools, open-data platforms and behavioral indicators to reinforce destination resilience and support the integration of sustainable tourism into wider territorial planning processes.

From a spatial planning perspective, the analysis underscores the value of climate-sensitive GIS mapping, visitor-flow and hazard monitoring as core components of urban and regional development strategies. Heritage centers such as Baku's Old city, Sheki and Ganja illustrate the need for planning approaches that balance accessibility with conservation imperatives and the protection of microclimatic conditions surrounding historic structures. Mountain tourism clusters including Gabala, Shahdag, Zagatala and Guba, as well as coastal zones such as Absheron and Lankaran, similarly benefit from risk-informed zoning, infrastructure optimization and corridor design that anticipate landslides, erosion, heat stress and other hazards. These approaches facilitate multi-criteria spatial evaluation and strengthen adaptive urban management [29].

International research further supports the conclusion that integrating ICT with urban planning methodologies enhances resilience outcomes. Evidence from Italy and Japan demonstrates that GIS-based hazard mapping improves the planning of urban infrastructure, emergency access routes and heritage-site protection. Studies from emerging economies show that digital resilience tools can be effective even in low-capacity environments, especially when built upon open-source geospatial platforms and mobile-based alert systems.

Azerbaijan's diverse climate zones, rapid ongoing digitalization and existing e-governance initiatives – such as ASAN service and national eGov platform – provide enabling conditions for expansion of ICT-supported spatial planning. Incorporating digital dashboards, hazard maps and mobility datasets into municipal and regional decision-making can help authorities optimize land use, safeguard natural and cultural assets and enhance visitor safety, while also supporting broader objectives related to sustainable economic development.

Several conceptual and methodological gaps identified in this study indicate the need for continued research. Future work should include empirical testing of ICT-based spatial planning tools in both urban and rural destinations, along with the integration of tourism and climate indicators into local authorities' GIS systems to support unified territorial management. Further inquiry into community-based digital platforms could strengthen participatory planning in environmentally and culturally sensitive areas, while the development and evaluation of predictive models may improve scenario planning for climate-indices risks. Strengthening interagency coordination – particularly between urban planners, tourism authorities and environmental institutions – remains essential to ensure that ICT tools are applied coherently and effectively.

In summary, ICT-enabled spatial planning frameworks represent a practical, scalable and context-sensitive pathway for improving the resilience of Azerbaijan's tourism

sector. Aligning digital tools with climate adaptation strategies and integrated urban planning can enhance the protection of natural and cultural heritage, strengthen territorial governance, and support sustainable growth across both urban and rural destinations.

## References


1. Dogru, T., Marchino, E.A, Bulut, U., Suess, C.: Climate change: vulnerability and resilience of tourism and the entire economy. Tourism Management Journal 72(3), 292–305 (2019). https://doi.org/10.1016/j.tourman.2018.12.010
2. State Tourism Agency of the Republic of Azerbaijan Homepage, https://tourism.gov.az/en , last accessed 2026/01/21.
3. AirClim Homepage, Northern Forests and Climate Change Project. https://www.air-clim.org/projects/northern-forests-and-climate-change/forests-azerbaijan , last accessed 2026/02/03.
4. United Nations Azerbaijan Homepage, https://azerbaijan.un.org/en , last accessed 2025/12/14
5. IGAZ Homepage, https://m.igaz.az/, last accessed 2026,01,26
6. Camera, G., Ismail, Y.: Adaptation of climate change: sustainable tourism development. ECOTOUR Ecotourism and Environment Conservation Journal 1(1), 1–8 (2024). https://doi.org/10.61511/ecotour.v1i1.2024.670.
7. Tsoggerel, G., Rehman, A, Rehman, K.U., Ullah, M.R, Zeeshan, M: Asymmetric effects of sustainable tourism and information and communication technology on environmental quality in China. Discover Sustainability 6, 1250 (2025). https://doi.org/10.1007/s43621-025-02132-6.
8. Lile, A., Kalluçi, E. Kaçurri, A: The impact of ICT on active tourism: a systematic literature review of innovation, smart tourism and sustainable development. International Journal of Digital Research 1(3), 46-52 (2025). https://doi.org/10.63711/ijdr.net20250305.
9. Bak, I., Barej-Kavzmarek, E., Sulikowski, P.: Impact of information and communication technologies on the tourism sector. European Research Studies Journal XXV (3), 595–606 (2022). https://www.um.edu.mt/library/oar/handle/123456789/104202
10. Jayasinghe, G.Y., Perera, T.A.N.T., Perrera, N.A.T.N., et al.: A systematic literature review on integrating disaster risk reduction (DRR) in sustainable tourism (SusT): conceptual framework for enhancing resilience and minimizing environmental impacts. International Journal of Disaster Risk Reduction 111 (2024) https://doi.org/10.1016/j.ijdrr.2024.104742
11. Almedia, D.S., Abreu, F.B, Boavida-Portugal, I.: Digital twins in tourism: a systematic literature review. Computers and Society (2025) https://doi.org/10.48550/arXiv.2502.00002
12. Bysrowska, M., Wigger, K., Liggett, D.: The use of information and communication technology (ICT) in Managing High Arctic tourism sites: a collective action perspective. Resources 6(3),33 (2017). https://doi.org/10.3390/resources6030033
13. Baydur, I.: Sustainable Development of regional tourism for economic resilience in Azerbaijan. Journal of Business and Economic Options 7(3), 10-18 (2024) https://resdojournals.com/index.php/jbeo/article/view/376
14. Musayeva, G.: Development prospects of the cultural tourism sector of Azerbaijan. Baku State University Journal of Earth Science & Environment 2(1), 19-29 (2025) https://doi.org/10.30546/209805.2025.2.1.120

15. Soltesova, M., Iannaccone, B., Strba, L., Sidor, C.: Application of GIS technologies in tourism planning and sustainable development: a case study of Gelnica. ISPRS International Journal of Geo-Information 14(3), 120 (2025). https://doi.org/10.3390/ijgi14030120
16. Li, J., Xu, L., Tang, L. et al.: Big data in tourism research: a literature review. Tourism management 68, 301-323 (2018) https://doi.org/10.1016/j.tourman.2018.03.009
17. Parra-Sanchez, D.T., Viviescas-Jaimes, P.A.: Application of digital technologies in nature-based tourism: a scientometric analysis. Journal of Digital Economy 3, 249-259 (2024) https://doi.org/10.1016/j.jdec.2025.05.003
18. Hagelsteen, M., Becker, P. Abrahamsson, M.: Troubling partnerships: perspectives from the receiving end of capacity development. International Journal of Disaster Risk reduction 59, 102231, (2021). https://doi.org/10.1016/j.ijdrr.2021.102231
19. Becken, S., Whittlesea, E., Loehr, J., Scott, D.: Tourism and climate change: evaluating the extent of policy integration. Journal of Sustainable Tourism 28(10), 1603–1624 (2020). https://www.tandfonline.com/doi/full/10.1080/09669582.2020.1745217
20. Zhang, Y., Tan, W.H., Zeng, Z.: Tourism demand forecasting bases on a hybrid temporal neutral network model for sustainable tourism. Sustainability 17(5), 2210 92025). https://doi.org/10.3390/su17052210
21. Munar, A.M., Jacobsen, J.K.S.: Motivations for sharing tourism experiences through social media. Tourism management 43, 46-54 (2014). https://doi.org/10.1016/j.tourman.2014.01.012
22. Birenboim, A., Bulis, Y., Omer, I.: A typology of tourism mobility apps. Tourism Management Perspectives 48, 101161 (2023). https://doi.org/10.1016/j.tmp.2023.101161
23. Kim, H., Koo, C., Chung, N.: The role pf mobility apps in memorable tourism experiences of Korean tourists: stress-coping theory perspective. Journal of Hospitality and Tourism Management 49, 548-557 (2021). https://doi.org/10.1016/j.jhtm.2021.11.003
24. Islam, T.: Applications of social media in the tourism industry: a review. SIESENSE Journal of Management 4(1), 59-68 (2021). https://doi.org/10.33215/sjom.v4i1.556
25. Lihmann, G., Eshaghi, M.S., Afshardoost, M., Moyle, B.D.: Weather adaptation strategies in tourism. Journal of Destination Marketing & Management 39, 101054 (2026). https://doi.org/10.1016/j.jdmm.2025.101054
26. Strong, S., Stewart, E.J., Espiner, S., Hanly, K.: The tourism adaptation classification (TAC) framework: an application to New Zealand's Glacier country. Frontiers in Human Dynamics 5 (2023). https://doi.org/10.3389/fhumd.2023.1130918
27. Leon, J., Winckler, P., Vicuna, M., et al.: Assessing the role of land-use planning in near future climate-driven scenarios in Chilean coastal cities. Sustainability 15(4), 3718 (2023). https://doi.org/10.3390/su15043718
28. Ahas, R. Aasa, A., Roose, A., et al.: Evaluating passive mobile positioning data for tourism surveys: an Estonian case study. Tourism management 29(3), 469-486 (2008). https://doi.org/10.1016/j.tourman.2007.05.014
29. Jimenez-Espada, M., Garcia, F.M.M., Gonzalez-Escobar, R.: Sustainability Indicators and GIS as Land-use planning instrument tools for urban model assessment. ISPRS International Journal of Geo-information 12(2), 42 (2023). https://doi.org/10.3390/ijgi12020042

# Digital Technologies and Their Impact on the Optimization of Business Processes in the Company: Focus on RPA

Kalen Meiirzhan Kanatuly

School of Information Technology and Engineering,
Kazakh-British Technical University, Almaty, Kazakhstan m_kalen@kbtu.kz

Abstract. Digital technologies are reshaping how companies optimize their business processes. In particular, Robotic Process Automation (RPA) has emerged as a powerful tool for automating repetitive tasks and improving process efficiency. This paper examines the impact of RPA on business process optimization in the context of an order processing workflow at a real company (LLP KazParts) in Kazakhstan. We model and implement an RPA solution that automates the end-to-end order pipeline—from online customer order entry through automatic creation of sales orders and payment postings in the 1C enterprise system to final shipment. A one-week experimental deployment is conducted to compare the RPA-driven process against the manual process in terms of processing speed, error rates, and throughput. The RPA implementation achieved significant performance gains, reducing average order handling time by over 80% and virtually eliminating data entry errors. However, several limitations were observed in handling complex exceptions and dynamic customer requirements. To address these shortcomings, we propose an improved hybrid RPA+AI model that combines rule-based automation with artificial intelligence components for greater flexibility. The results demonstrate the substantial benefits of RPA for business process optimization in the company while highlighting the value of integrating AI to handle process variability.

Keywords: Robotic Process Automation · Business Process Optimization · Digital Transformation · Order Processing · Intelligent Automation · 1C Enterprise

## 1 Introduction

Digital transformation is driving organizations to leverage advanced technologies for optimizing business processes. Among these technologies, *Robotic Process Automation* (RPA) has gained prominence as a means to automate repetitive, rule-based tasks in information systems. RPA involves software robots that emulate human interactions with user interfaces to perform tasks, effectively acting as a scalable digital workforce [2]. By offloading mundane tasks to software bots, companies can

achieve higher operational efficiency, reduced error rates, and improved processing speed [1]. Studies report that RPA implementations lead to significant cost savings and productivity gains across various industries [1,4,5]. For instance, in banking and insurance sectors, RPA has shortened processing times and lowered operational costs by automating high-volume transactions [4]. Real-world case studies further exemplify these benefits: Telefónica O2 deployed over 160 RPA bots to handle about half a million transactions per month, achieving a 650–800% return on investment and reducing some process cycle times from days to minutes [7,6]. Such outcomes have led researchers to identify RPA as a key lever for process transformation in service operations [19] and an enabler of digital transformation initiatives in organizations [3,17].

Despite its potential, successful RPA adoption is not trivial. Organizations must carefully plan which processes to automate and address several implementation challenges [1,5]. One critical factor is selecting suitable candidate tasks for RPA that are rule-based and stable [20]. Poor process selection or attempting to automate highly complex, exception-laden tasks can lead to suboptimal outcomes. Another challenge lies in integrating RPA with legacy enterprise systems; issues such as limited APIs or changing user interfaces can cause reliability problems [8]. Additionally, organizational and human factors play a significant role. Introducing software robots often requires managing employee concerns and change aversion, as well as redesigning workflows to accommodate human–bot collaboration [5,8]. Studies have identified technical barriers and organizational barriers that can hinder RPA projects [5]. Ensuring proper governance and change management is thus essential for realizing the full benefits of RPA [1]. Moreover, RPA should not be viewed as a one-time solution; continuous monitoring and improvement of bots are required to adapt to evolving business processes [16].

In the research community, there is a growing body of work examining RPA applications and outcomes. Recent literature reviews summarize contemporary RPA research themes, benefits, and challenges [10]. These include case studies in domains such as finance, healthcare, and supply chain, which generally report improved efficiency and accuracy from RPA deployments. In the supply chain context, RPA has been used to streamline order fulfillment and inventory management processes [9]. However, there is limited academic literature focusing on RPA implementations in Central Asian companies or on integration with regionspecific enterprise software. This study aims to address this gap by analyzing an RPA deployment at a company in Kazakhstan that uses the local 1C enterprise platform for trading and accounting.

## 2 Methodology

### 2.1 RPA Workflow Design

The target process for automation is the customer order processing workflow, which spans multiple systems, namely the web storefront and the internal 1C enterprise

applications. We first mapped the *as-is* manual process to identify the sequence of activities and decision points. In the manual workflow, a sales operator periodically checks for new online orders, then logs into the 1C:Trade Management system to input each order's details, including customer information, ordered items, quantities, and prices. Next, the operator switches to the 1C:Accounting module to record the payment associated with the order or mark it as cash on delivery. Finally, an invoice and shipping document are generated in 1C, and the warehouse is notified to pick, pack, and dispatch the items.

Figure 1 illustrates the *to-be* automated workflow after introducing RPA. Instead of a human operator performing data transfer and data entry, an RPA bot executes these actions across the web platform and the 1C systems.

In the automated workflow, the RPA bot monitors the e-commerce platform for newly placed orders by querying the order database or application programming interface. Once a new order is detected, the bot extracts all relevant data such as the order identifier, customer details, shipping address, item codes, quantities, and payment status. Using this information, the bot logs into the 1C:Trade Management system through its graphical user interface and navigates the required menus and forms to create a new sales order. All mandatory fields are populated automatically in the same manner as a human operator would, but significantly faster.

After creating the sales order, the bot proceeds with payment processing. For prepaid orders, the bot switches to the 1C:Accounting system and records the corresponding financial transaction. If the order is marked as cash on delivery or another deferred payment method, the bot skips this step and flags the order accordingly. Subsequently, the bot generates the invoice and shipping documentation in 1C and reserves the required stock for dispatch. As a final step, the warehouse team is notified automatically, for example via email or a shared task list, that the order is ready for shipment.

To ensure reliability, the automation was implemented using stable user interface element selectors and predefined delays to accommodate screen loading times. The bot was developed on a commercial RPA platform compatible with the Windows-based 1C environment. Although direct API or database integration is sometimes possible, user interface automation was selected for this project due to its flexibility and minimal impact on the existing IT infrastructure. Basic exception handling mechanisms were implemented; for example, if a product code was not found in the 1C catalog, the bot logged the incident and skipped the order for manual review.

## 2.2 Algorithm and Pseudocode

The logical flow of the RPA bot was formalized using pseudocode to guide implementation and ensure reproducibility. Algorithm 1 presents the core logic of the automated order processing procedure.

As shown in Algorithm 1, the bot operates in a continuous loop and processes orders sequentially from start to finish. Each iteration handles a single order, ensuring data consistency across systems. In practice, the bot can be configured

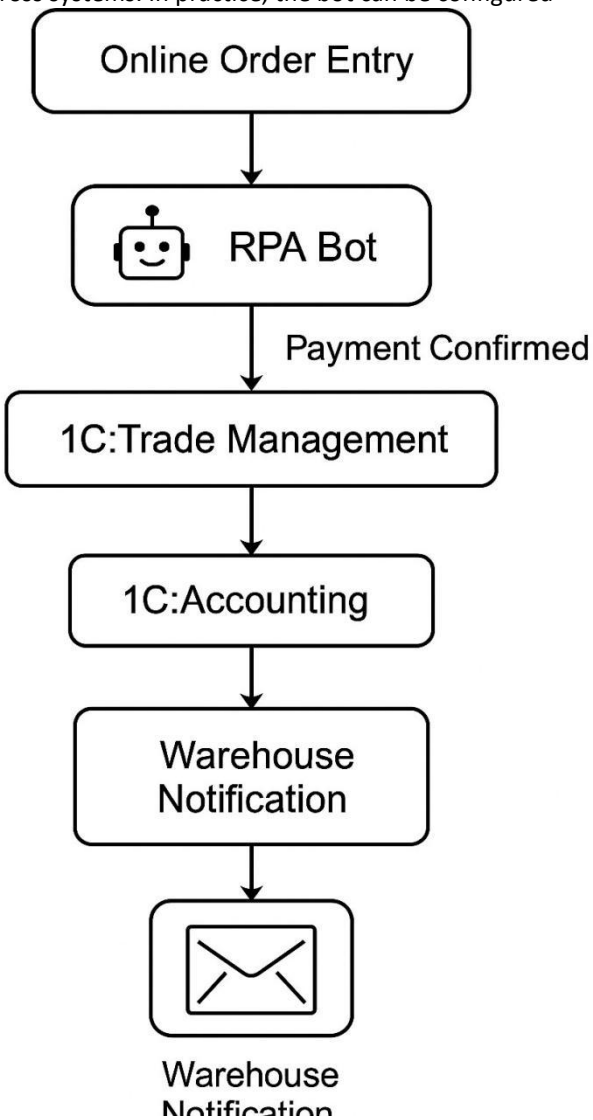



Fig.1. Overview of the RPA-based order processing workflow, from online order entry to update in 1C and warehouse notification.

Algorithm 1 RPA Order Processing Procedure

1: Initialize RPA bot and establish connections to the web order system and 1C applications
2: while new online orders are available do
3: Retrieve order data from the web platform

```
4:          Launch or activate 1C:Trade Management and authenticate
5:          Create a new sales order using the retrieved order information
6:         if order payment is confirmed then
7:            Launch or switch to 1C:Accounting
8:             Record the payment transaction in the accounting system
9:       end if
10:        Generate invoice and shipping documents in 1C
11:        Notify the warehouse that the order is ready for shipment
12:       Mark the order as processed in the web system
13: end while
```

to run at fixed time intervals or be triggered in near real-time when new orders are created.

### 2.3 Experimental Setup and Metrics

To assess the effectiveness of the RPA implementation, a pilot deployment was conducted and compared against the baseline manual process. The evaluation period lasted seven consecutive days and covered a representative volume of online orders.

For the baseline scenario, performance data were collected from historical records and direct observations of manual processing. The following metrics were measured:

- Processing time per order, defined as the elapsed time from online order placement to completion of data entry and documentation in the 1C system.
- Throughput, measured as the number of orders processed per hour or per working day.
- Data entry error rate, defined as the proportion of orders requiring correction due to incorrect or incomplete information.

During the RPA pilot, the same metrics were captured for the automated process:

- Automated processing time, measured from order creation on the website to completion of all automated actions in 1C.
- Throughput of the RPA bot, expressed as the number of orders processed per hour.
- Exception rate, defined as the percentage of orders that could not be fully processed by the bot and required human intervention.

To quantify the efficiency gain achieved through automation, we used the following metric:

$$\eta = \frac{T_{\text{manual}} - T_{\text{RPA}}}{} \times 100\%, \quad (1)$$

$T_{\text{manual}}$

where $T_{\text{manual}}$ denotes the average processing time per order in the manual scenario, and $T_{\text{RPA}}$ denotes the average processing time per order under RPA automation. The resulting value $\eta$ represents the relative reduction in processing time achieved through automation.

The collected data were aggregated and analyzed to compare the manual and automated processes. The next section presents the experimental results and discusses the observed performance differences.

## 3 Results and Discussion

During the seven-day pilot deployment, the RPA bot processed a total of 320 online orders, corresponding to an average of approximately 45 orders per day. The evaluation compared the performance of the manual order processing workflow with the RPA-automated workflow. Table 1 summarizes the observed performance differences in terms of processing time, throughput, and error or exception rates.

*Processing Time and Throughput* In the manual workflow, a sales operator required approximately 8–10 minutes on average to complete the processing of a single order. This corresponds to a throughput of roughly six orders per hour per employee. In contrast, the RPA bot completed the same set of activities in an average of 1.5 minutes per order. Due to its ability to operate continuously and without interruptions, a single bot instance achieved an effective throughput of approximately 40 orders per hour.

Using the efficiency improvement metric defined in Eq. (1), the reduction in average processing time per order reached approximately 80%. This result indicates that the automated process was about five times faster than the manual process. Such performance gains are consistent with findings reported in previous studies, where RPA deployments led to time savings of 70–80% in transactionoriented processes [4,10]. From an operational perspective, the faster processing time enabled near-immediate order confirmation and earlier initiation of shipment activities, which previously could be delayed by several hours or even until the next business day.

*Accuracy and Error Rates* Manual data entry was found to introduce occasional errors in the baseline process. During the observation period, approximately 3% of manually processed orders required corrections due to incorrect quantities, item codes, or customer information. These errors resulted in additional rework and, in some cases, delayed shipment.

The RPA-based process eliminated manual transcription and achieved zero data entry errors for all orders processed successfully by the bot. Since order data

PERFORMANCEE OF MANUAL PROCESS VS. RPA AUTOMATION OVER 7 DAYS

| ?erformance Metric | Manual Process | RPA Automatic |
|---|---|---|
| Average Processing Time per Order | 9.0 minutes | 1.5 minut |
| Daily Throughput | 45 orders | 320 orde |
| Error/exception Rate | 4.7% | 0.6% |
| Error–r(axsed naduction | 4.4% | — |

Table 1. Performance comparison between the manual process and the RPAautomated process over seven days.

were transferred directly from the web platform to the 1C systems, the risk of human error was effectively removed. This outcome aligns with widely reported benefits of RPA in improving data accuracy and process reliability [1,2]. However, the automated process experienced a small number of exceptions. Specifically, five orders, corresponding to approximately 1.6% of the total, could not be fully processed by the bot and required manual intervention. These exceptions were caused by conditions such as missing product records or unexpected validation requirements in the 1C system. While such cases do not represent errors in the traditional sense, they indicate situations where rule-based automation was insufficient.

*Resource Utilization* The introduction of RPA significantly altered the allocation of human resources. Prior to automation, two employees were responsible for handling online orders on a daily basis. During the pilot, the RPA bot performed the majority of routine processing tasks, allowing these employees to focus on higher-value activities such as customer communication, exception handling, and coordination with suppliers. One employee remained responsible for monitoring the bot and resolving exceptional cases, but the overall manual workload was substantially reduced.

From a managerial perspective, this shift demonstrates how RPA can improve labor productivity and operational scalability. The company could process a larger

volume of orders without increasing staffing levels or could maintain current volumes with reduced overtime. Similar effects have been reported in other RPA case studies, where organizations achieved measurable cost savings and improved utilization of human expertise [5,6].

*Challenges and Shortcomings* Despite the clear benefits, several limitations of the RPA implementation were identified. The rule-based nature of the bot restricted its ability to handle unstructured or unexpected situations. For example, customer orders containing special instructions or requests outside the predefined process flow were not interpreted by the bot and required later human review. In addition, when items were out of stock, the bot was unable to make contextual decisions such as splitting shipments or proposing alternative products.

These findings reflect a broader limitation of traditional RPA systems, which excel at standardized, repetitive tasks but lack the cognitive capabilities required for judgment-intensive activities [10,16]. Furthermore, the reliance on user interface automation introduced maintainability concerns. Minor changes to the 1C interface, such as updated layouts or renamed fields, could potentially disrupt the bot's operation and require reconfiguration. This issue has been widely recognized in the literature as a key challenge in sustaining RPA solutions over time [8].

*Comparison with Existing Literature* Overall, the observed results are consistent with prior research on RPA effectiveness. The magnitude of time savings and error reduction aligns with reported outcomes in financial services, procurement, and logistics processes [7,4]. At the same time, the presence of exceptions and the need for ongoing human oversight mirror findings from multiple-case studies, which emphasize that RPA alone does not fully eliminate the need for human involvement [5]. These insights support the growing consensus that RPA should be viewed as a foundational automation technology that benefits from integration with complementary approaches such as artificial intelligence. In the following section, we discuss how a hybrid RPA+AI model can address the identified limitations and further enhance process performance.

## 4 Conclusion

This study demonstrated how robotic process automation can significantly optimize a company's business process, using the example of order processing at a mid-sized enterprise. By deploying an RPA bot to automate the transfer of online orders into the 1C enterprise system, we achieved major improvements in efficiency (over 80% reduction in processing time per order) and accuracy (elimination of manual entry errors). The practical implication for the company, LLP KazParts, is faster order fulfillment and improved operational scalability. Customers benefit from quicker order confirmations and deliveries, while the company can handle higher order volumes without proportional increases in labor. These results underscore the positive impact

of digital technologies like RPA on process performance and support the broader narrative of RPA as a driver of process streamlining and cost savings in organizations [1,5].

At the same time, our findings highlight that RPA, in its traditional form, has limitations when confronted with complex or dynamic scenarios. The pilot revealed that exceptions (such as special customer requests or unforeseen process variations) cannot be effectively managed by a static rule-based bot. This is a known challenge with RPA implementations — bots are only as smart as the rules they are given, and they lack the human ability to adapt or understand context beyond those rules [16,10]. As business processes are rarely 100% stable or predictable, it is important to plan for these exceptions. In the current implementation, we addressed them by having humans handle out-of-scope cases, but this partially reintroduces manual work and slows the end-to-end process for those instances.

To further improve and generalize the solution, we propose moving toward a hybrid RPA+AI model. In such a model, the RPA bot would continue to handle the deterministic, high-volume tasks, while artificial intelligence components are integrated to manage more complex decision points. For example, a natural language processing (NLP) module could be used to read and interpret special instructions or customer notes attached to orders, allowing the system to take appropriate actions or alerts automatically. Machine learning models could predict and flag orders that are likely to face issues (e.g., an item likely out-ofstock or a potentially fraudulent order) so that they can be routed for special handling before errors occur. An intelligent agent could also dynamically decide on fallback actions (like suggesting an alternative item for an out-of-stock product, or automatically contacting the customer for clarification) instead of simply halting as the current bot does. Prior research advocates this kind of intelligent automation approach: blending RPA with AI techniques can overcome the rigidity of standalone RPA and extend automation to semi-structured tasks [4,15]. In the literature, this convergence is sometimes referred to as “hyperautomation,” where RPA, AI, and other digital tools collectively enable end-to-end automation of more complex processes [15,11].

For the company in question, implementing a hybrid RPA+AI solution could mean incorporating an AI-driven decision support system within the order processing pipeline. For instance, an AI module could interface with the inventory system to check for stock levels and automatically decide to split an order shipment if an item is unavailable (something a basic RPA bot would not do on its own). Such integration of AI would require additional development and training data, but it promises to further reduce the need for human intervention and increase the resilience of the automated process. Indeed, early experiments in other contexts have shown that adding AI (such as computer vision or predictive analytics) to RPA can enable automation of tasks that involve unstructured data or complex rules, which were previously beyond RPA’s scope [14,13].

In conclusion, the implementation of RPA at LLP KazParts clearly showcased the power of digital automation in optimizing business processes. The project achieved its

immediate goal of accelerating the order-to-shipment cycle and minimizing errors, thus providing value to both the company and its customers. It also offered insights into how RPA needs to be managed — with proper process selection, ongoing maintenance, and complementary strategies for handling what RPA alone cannot. The proposed hybrid RPA+AI model is a logical next step towards more intelligent process automation. Future work will involve developing and testing this hybrid approach, for example by integrating an AI-based exception handler into the current RPA workflow and evaluating its performance over a longer term. We expect that such an enhanced system would further reduce the exception rate and could adapt better to changes in business conditions or customer behavior. The lessons from this case study contribute to the growing knowledge base on how organizations can practically leverage RPA and related digital technologies to achieve process excellence in the era of digital transformation.

## Acknowledgment

We would like to thank the management and IT department of LLP KazParts for their support and collaboration in this project. We also acknowledge the process owners and employees who provided valuable insights into the workflow and assisted during the RPA pilot deployment. This work was supported in part by the Company's Digital Innovation Initiative 2025.

## References

1. N. Katiyar *et al.*, "Robotic Process Automation (RPA) in business operations: opportunities and implementation strategies," *Educational Administration: Theory and Practice*, vol. 30, no. 1, pp. 1137–1144, 2024.
2. S. Madakam, R. M. Holmukhe, and D. Jaiswal, "The future digital workforce: Robotic Process Automation (RPA)," *J. Inf. Technol. Case Appl. Res.*, vol. 21, no. 1, pp. 1–11, 2019.
3. P. Hofmann, C. Samp, and N. Urbach, "Robotic process automation as an enablerof digital transformation: an overview of benefits and challenges," in *Proc. 27th Eur. Conf. Inf. Syst. (ECIS)*, 2020.
4. B. T. Harris and O. N. Martinez, "The expanding role of robotic process automationin improving business process management," 2025.
5. C. Flechsig, F. Anslinger, and R. Lasch, "Robotic process automation in purchasingand supply management: A multiple case study on potentials, barriers, and implementation," *J. Purchasing Supply Manag.*, vol. 28, no. 1, p. 100718, 2022.
6. M. Lacity, L. Willcocks, and A. Craig, "Robotic process automation at TelefónicaO2," *MIS Q. Executive*, vol. 15, no. 1, pp. 21–35, 2016.
7. R. Jayasekhar, D. Jayasinghe, and J. Ahmed, "Adopting robotic process automationto unlock business value," *J. Financial Transformation*, vol. 46, pp. 98–109, 2018.
8. H. Putra and M. Er, "The changing function of business process management inrelation to digital transformation," 2024.

9. A. Shamsuzzoha and S. Pelkonen, “A robotic process automation model for orderhandling optimization in supply chain management,” *Supply Chain Analytics*, vol. 9, no. 4, p. 100102, 2025.
10. R. Syed *et al.*, “Robotic process automation: contemporary themes and challenges,” *Computers in Industry*, vol. 115, Art. no. 103162, 2020.
11. J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, “Robotic process automation andartificial intelligence in Industry 4.0 – a literature review,” *Procedia Comput. Sci.*, vol. 181, pp. 51–58, 2021.
12. S. Aguirre and A. Rodriguez, “Automation of a business process using roboticprocess automation: a case study,” in *Proc. Workshop Eng. Applications (WEA)*, 2017, pp. 1–6.
13. M. Rosemann, “Harnessing artificial intelligence for next-generation business process management,” 2024.
14. V. R. Zambrano-Burgos, J. D. Zambrano-Mieles, and D. M. Cevallos, “El rol de lainteligencia artificial en la transformación digital de la ingeniería industrial, con un enfoque en la automatización y la gestión de la cadena de suministro,” Feb. 2025.
15. A. Royhan, H. S. Ahluwalia, A. Ramachandran, and M. B. Ushrani, “Structuring intelligent automation: a hyperautomation taxonomy,” in *Proc. 56th Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2023.
16. S. Laumer, C. Maier, A. Eckhardt, and T. Weitzel, “The trend is your friend: howBPM practitioners perceive and cope with robotic process automation,” *Computers in Human Behavior*, vol. 99, pp. 168–177, 2019.
17. W. M. P. van der Aalst, M. Bichler, and A. Heinzl, “Robotic process automation,”*Bus. Inf. Syst. Eng.*, vol. 60, no. 4, pp. 269–272, 2018.
18. A. Tayab and Y. Li, “The contribution of robotic process automation in enhancingbusiness performance in warehouse management,” *IEEE Access*, 2024.
19. L. Willcocks, M. Lacity, and A. Craig, “Robotic process automation: the nexttransformation lever for shared services,” LSE Working Paper, no. 15/01, 2017.
20. Y. Kalini˘cenko, O. Turetken, W. van Genuchten, and H. Weigand, “Selecting candidate tasks for robotic process automation,” in *Proc. 35th Bled eConf.*, 2022, pp. 345–360.

# Enhancing Learning Effectiveness and Satisfaction through Game-Based Learning in Higher Education

Tam Tran Luu[0009-0008-4443-064X] and Luan Trong Nguyen[0000-0002-3489-1628]

FPT University, Can Tho Campus, Vietnam
tranltcs201170@gmail.com

**Abstract.** In modern education today, integrating advanced technologies into education is becoming increasingly common, and researching the impact of technology on learning is essential. This study will focus on exploring aspects surrounding learning motivation and the effectiveness that technology brings through understanding and analyzing the game-based learning (GBL) design framework. In higher education, teaching and learning methods that provide students with the most multifaceted opportunities for self-development require educators and institutions to focus on deeper development of flexible and unconstrained curriculum and teaching design frameworks. This allows students to feel comfortable and unpressured by their studies. With the GBL model, this is considered the basis for combining it with traditional teaching methods to create learning motivation and participation, thereby training the ability to adapt and update trends quickly and effectively. Despite growing interest, there is still limited empirical evidence on how design factors can influence learning outcomes and student satisfaction with their experience of the model throughout their studies. Insights from qualitative research have emphasized the importance of a clear alignment between game mechanics and learning objectives. The results of the study will partially clarify the factors and thus provide the correct directions for learning thru the GBL model and applying it in practice, applying existing theories and models to enhance the superiority of the GBL model, and providing information most relevant to the needs of students and educators when they need additional input to strengthen and develop the model to become increasingly appropriate.

**Keywords:** Game-Based Learning (GBL), Learning Effectiveness, Student Satisfaction, Motivation and Engagement, Educational Technology.

## 1 Introduction

21st century education, technology is indispensable, which is the driving force for the development of teaching and learning methods[1]. Previous Traditional teaching methods are valuable and practiced over the years, but are sometimes limited in maintaining student motivation and raising the level of higher-order thinking. Later, educators and researchers introduced innovative methods using technology to enhance learning motivation, and aid in improving knowledge retention[2]. One of them is the game-based learning (GBL) method, which has advanced as a breakthrough in the education industry, a tool that knows how to conceptualize many new aspects that other learning methods have not been able to do before. Leverage state-of-the-art technologies that engage students so that they feel interested in learning and achieving their goals during their studies[3], a number of studies have explored the process of introducing AI technology or game-based learning to research and evaluate how to develop how this application will be optimized and most suitable for students[4]. Research in many fields such as

learning science, mathematics and professional skills training. Studies have highlighted the effectiveness and potential of GBL in creating the appeal of game-based learning to enhance learner engagement[5], promote proactive problem-solving, and encourage interaction. However, there are still many challenges in ensuring that effective games are built to suit each student, and there is a need for an intelligent model design approach that does not let students focus solely on playing and forget about the learning task[6]. Without a clear design framework, there will be many problems in educators systematically integrating GBL into the curriculum[7]. While there has been a lot of serious gamification and gamification research in the past, little research has focused on how to develop and create comprehensive design frameworks for GBL in higher education settings. Models that are commonly applied and familiar to students, such as the ARCS model or the Catalysis framework, are an important basis for providing valuable insights into the effects on motivation and engagement, but they still lack direct application to teaching model designs. The gap for this issue underscores the need for new creativity in terms of robust and practical theoretical frameworks, in order to provide innovative ideas and the right direction for educators and instructional model designers to help develop worthwhile GBL experiences[8]. The purpose of this study is to propose a Game-Based Learning Design Framework for students, the main content of which is to explore and identify different modes of coordination based on the key principles of learning theory, motivation to participate and experiential learning[9]. The results can contribute to the wealth of information educators can take in two important directions. First, it refers to a structural framework – called the 5D-GBL Framework – which synthesizes educational theories with game design principles, providing systematic approaches to implementing GBL in the most optimal way[10]. Simultaneously, it provides suggestions based on its findings and will point out potential directions that educators can take to integrate GBL into their teaching, which will be a boost to innovation in a more diverse and effective learning environment. This study aims to answer three major questions:

(1) What are the essential elements of a high-quality GBL design?

(2) How does GBL influence learners' motivation and academic outcomes?

(3) How can a model framework be developed that can support educators in addressing all issues in different contexts?

If there's a way to address these questions, the research aims to contribute both theoretically and practically to the current field of digital technology education.

## 2 Literature review

### 2.1 Definition of GBL and distinguishing concepts

Game-Based Learning (GBL) is defined as the use of game elements and mechanics in learning environments, with its primary focus being learning and serving as a tool for students to achieve their goals. While GBL prioritizes the development of instructional designs, gamification is merely seen as a model to make lessons more engaging through elements like points, badges, or leaderboards[11]. GBL places games at the center of

the learning process[12]. Similarly, serious games are designed for purposes beyond entertainment but still need to closely follow the learning process, although they don't always align perfectly with the curriculum[13]. Therefore, GBL could be a game-changer in a new approach to differentiated instruction, requiring upfront investment and knowing how to intentionally integrate game design and instructional design[14].

### 2.2 Fundamental Theoretical Basis and Related Models

The application of the GBL model can be explained by various educational theories[15]. From a constructivist perspective, learners actively learn and draw meaning from the lesson through experience, which aligns with learning through games[16]. Kolb's experiential learning theory emphasizes drawing meaning and knowledge through concrete experience, which are also elements present in GBL[17]. Regarding learning motivation, Self-Determination Theory (Deci & Ryan) identifies three factors that drive intrinsic motivation: autonomy, competence, and social connection[18]. Additionally, Flow Theory (Csikszentmihalyi) explains the phenomenon of "immersion" as occurring when the learner is and has been in a state of balance between task and ability[19]. These theoretical foundations strengthen the scientific basis for GBL in both cognitive and emotional aspects of learning experiences through theories[20]. Educators have seen the potential and effectiveness of theories when applying games to learning models. Many instructional design models and motivational methods have been applied in GBL research. For example, Keller's ARCS model (Attention, Relevance, Confidence, Satisfaction) focuses on developing learning motivation[21]. However, this model focuses on the motivational aspect of learners but doesn't address the crucial need for a clear strategy to ensure the game is designed to directly align with teaching objectives and outcome assessment[22]. Therefore, they also provide useful ideas but cannot yet become the main GBL design framework, but this will be important information for comprehensively incorporating and developing the GBL model.

### 2.3 Experimental Information and Research Gaps

Many previous experimental studies have demonstrated the effectiveness of GBL in various fields. In language education, games help improve quick vocabulary recall and increase English communication skills. In the STEM field, GBL is integrated through simulations to solve problems and support the generation of ideas in terms of logical thinking. In specialized skills, especially for specific jobs like medicine and business, games provide scenarios for learners to practice problem-solving before applying it in real-world situations, thereby limiting risks when making decisions in practice[23]. However, studies also point out challenges such as high operating costs for the model and complex technical requirements for the machinery. The research results show that GBL offers many benefits in terms of motivation to participate and effectiveness to learn, but there is still a lack of a systematic design framework that allows instructors and educational designers to easily incorporate it into lessons so that students have the

best conditions for development through the model[24]. This gap sets the stage for the current research, which aims to propose a comprehensive Game-Based Learning Design Framework that is both theoretically sound and familiar to students' practical experiences[25].

# 3 Methodology

## 3.1 Research Design

The research design is based on a conceptual framework for Game-Based Learning (GBL). As illustrated in Figure 1, the proposed model hypothesizes the relationships between three independent variables—Motivation (MOT), Engagement (ENG), and Usability (USA)and two dependent variables, namely Learning Performance (LEA) and Satisfaction (SAT). Independent variable: Motivation (MOT) This is intrinsic motivation that encourages learners to participate regularly in GBL activities. Engagement (ENG) defines the level of involvement in terms of connection with people and awareness of learning. Usability (United States) refers to the ease of use and accessibility of the GBL platform. Intermediate variable: Learning effectiveness (LEA) is conceptualized as the improvement in knowledge and experiences discovered from GBL experiences. Dependent variable: Satisfaction (SAT) indicates the overall satisfaction of learners with the learning environment using the GBL model.
Based on the information and theoretical framework, we will formulate the following hypotheses:

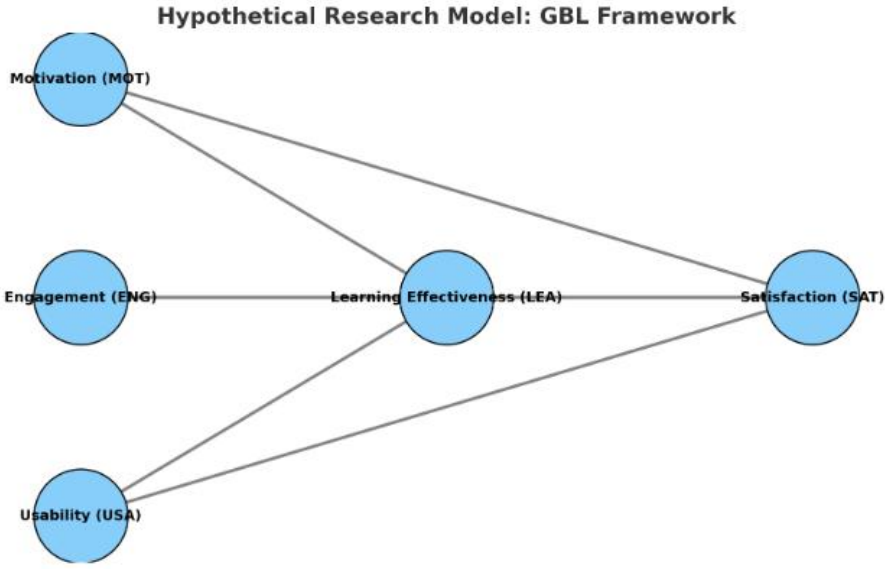



**Fig. 1.** Proposed Research Model for the GBL Framework

The study includes 7 hypotheses, which are: H1: Motivation positively influences Learning Effectiveness, H2: Motivation positively influences Satisfaction, H3: Engagement positively influences Learning Effectiveness, H4: Engagement positively influences Satisfaction, H5: Usability positively influences Learning Effectiveness, H6: Usability positively influences Satisfaction, H7: Learning Effectiveness positively influences Satisfaction.

**Table 1.** In-depth questions

| Project | Code | Question |
|---|---|---|
| Motivation & Engagement | Q1 | A recent experience when you learned through games (GBL): the situation, the tasks, the things that made you feel interested and helped increase your motivation to participate in your learning? |
| Game Mechanics & Usability | Q2 | In your opinion, which elements of the game design (e.g., content, challenges, reward system, interface/navigation, teamwork, integration with learning content) help you understand the material best? And explain specifically with examples. |
| Transferability & Learning Effectiveness | Q3 | Provide a specific example of when you applied GBL (or didn't) and were able to use the knowledge/skills learned from the game in an exercise or real-life situation. If you haven't been able to practice yet, what do you think the reason is? |
| Satisfaction | Q4 | If you were given the opportunity to participate in a significant improvement process to increase the effectiveness and satisfaction of using GBL in the subject, what would you suggest? |

### 3.2 Participants

This survey-based study focuses on university students who have or are currently using the GBL model in their learning process, have experience using AI technology, and learn through games. By examining students' perspectives and real-world experiences, educators will gain valuable feedback on how students have experienced and perceived the GBL model when applying it to their learning process, and what impact it has had. This will provide valuable insights into learning opportunities and highlight challenges that need to be addressed. An online survey on the variables in the GBL model was conducted via Google Forms and yielded 150 valid responses (n=150). The demographic information of the participants is presented in (Table 1):

**Table 2.** Results of frequency analysis on demographic variable

| | | Frequency | Percentage |
|---|---|---|---|
| Gender | Male | 71 | 47.3% |
| | Female | 79 | 52.7% |
| Student | Freshman | 41 | 27.3% |
| | Sophomore | 35 | 23.3% |
| | Junior | 35 | 23.3% |
| | Senior | 39 | 26% |
| Major | Information Technology | 68 | 45.3% |
| | Business Administration | 44 | 29.3% |
| | Communication Technology | 19 | 12.7% |
| | Languages. | 19 | 12.7% |
| GBL usage frequency | Daily | 43 | 28.7% |
| | Weekly | 68 | 45.3% |
| | Monthly | 27 | 18% |
| | Never | 12 | 8% |
| Game-Based Learning (GBL) acceptance | Yes | 137 | 91.3% |
| | No | 13 | 8.7% |

### 3.3 Instrument

Two main tools have been conducted: Questionnaire scale: a questionnaire-based survey using the 5-point Likert scale that has been conducted to measure students' perceptions of motivation, participation, ease of use that impact learning outcomes, and their satisfaction with experiencing the model in the real world of the learning process. The questions were explored and identified based on previous model theory experiments when studying GBL to ensure high reliability and accuracy of the core content of the model.

In-depth interviews: Conduct in-depth interviews with information technology majors and faculty to gain more specific information about the educational opportunities, challenges, and feasibility of applying this model to the implementation of the formal curriculum framework. Specifically, the interview information includes

open-ended questions related to the design creation process, and receives student feedback and support from faculty and the school.

### 3.4 Data Collection Procedure and Analysis

This study was conducted using a mixed methods approach, combining quantitative and qualitative research. The quantitative survey was conducted through Google Forms with 20 quantitative questions to assess the level of the variables, along with 4 in-depth survey questions aimed at understanding the variables that directly impact learners' perceptions of applying the GBL model. The learner data collection process took place in September 2025, yielding 150 student responses through quantitative survey questions and variables. We proceeded to the data analysis stage, including Cronbach's alpha reliability analysis, t-tests, mean statistics, frequency statistics, one-way ANOVA, and regression. The range of values for the analysis variables is determined and evaluated using these statistical analysis methods, which is a way to describe and present results from the data in an accessible and comprehensive manner. The descriptive statistical indicators used have been explained in accordance with the research objectives. The average scores of the variables under consideration are based on the Likert scale, a common measurement method in research used to assess abstract variables related to behavior and attitudes, to draw conclusions.

## 4 Results and discussion

### 4.1 Survey results

**Table 3.** Cronbach's Alpha, Mean, and Standard Deviation

| Criteria | N | Cronbach's alpha | Mean | SD |
|---|---|---|---|---|
| Motivation (MOT) | 150 | 0.796 | 4.3000 | .65053 |
| Engagement (ENG) | 150 | 0.795 | 4.2689 | .64932 |
| Usability of GBL Design (USA) | 150 | 0.852 | 4.2689 | .72937 |

Table 3 shows the results of the variable analysis with Cronbach's alpha reliability co-efficient, the mean and standard deviation values of the three observed variables. The Cronbach's alpha values for all variables on the scale are above the threshold of 0.70, indicating good internal consistency among the variables and achieving an accurate result of (MOT =. 796; ENG =. 795; USA =. 852). Among the analysis variables, the usability of the USA design of the GBL design showed the highest reliability ($\alpha$ = 0.852). The average scores for all variables are relatively high, ranging from 4.27 to

4.30 on a 5-point observation scale. This indicates that students generally have a positive view of the factors influencing motivation, engagement, and usability in game-based learning. The standard deviation ranged from 0.65 to 0.73, reflecting a moderate variation with no significant difference between the responses. The analysis results show that the variables measuring motivation, engagement, and usability for the model are reliable and suitable for this study. A special feature is that when the usability reaches the highest level of reliability and consistency, this serves as the basis for evaluating its robustness and suitability in assessing students' perspectives on GBL design. The average value across all factors is relatively high, indicating that students have a positive perception of the GBL model in promoting motivation and engagement, while also highlighting its user-friendliness. These results are consistent with previous studies, not only emphasizing that usability is a highly influential factor in learner satisfaction and effectiveness in the learning environment but also demonstrating adaptability and agreement when technology is supported. Motivation and engagement were positively evaluated, with high reliability and average values for the core variables, indicating their potential role in shaping learners' overall experience with GBL.

**Table 4.** Independent Samples T-Test and One-way ANOVA

| Criteria | N | Sig. (Independent sample T-test) | Sig. (2-tailed) Equal variances assumed | Sig. (Test of Homogeneity of Variances) | Sig. (ANOVA) |
|---|---|---|---|---|---|
| Motivation (MOT) | 150 | .837 | .732 | .522 | .733 |
| Engagement (ENG) | 150 | .461 | .820 | .348 | .830 |
| Usability of GBL Design (USA) | 150 | .317 | .343 | .733 | .768 |

Independent t-tests were analyzed to obtain results on motivation, engagement, and usability of the GBL design to examine if there were significant differences between male and female students. The results show no significant statistical differences in outcomes based on the two-variable survey samples across the three observed variables (Motivation: p =. 837; participation: p =. 461; Usability: p = .317), indicating that both gender

variables have relatively similar perceptions of GBL. The test for homogeneity of variance is not statistically significant, and the ability to confirm the fit is the same. One-way analysis of variance (ANOVA) was performed to examine and evaluate whether students' academic majors influenced their perceptions when three variables related to the GBL model were considered. The results showed no significant difference between the study fields regarding motivation (p =. 733), participation (p =. 830) or usability (p =. 768). These findings reflect students' consistent experiences with GBL, which are not influenced or dependent on their major. The analysis process examining differences by gender or major shows that students' overall perceptions of GBL are positive and that it is accepted and shared among students, regardless of demographic factors. This finding suggests that GBL values are designed to be effective and can provide an inclusive and diverse learning environment for various groups of learners. Besides, identifying usability and motivation as fundamental factors of GBL, it can overcome personal factors such as gender or academic expertise. Previous studies also suggest that the effectiveness and suitability of GBL depend heavily on the process of conceptualizing the design framework and incorporating educational and academic elements, rather than focusing on evaluating demographic variables.

**Table 5.** Results of regression model

| Criteria | N | R Square | Adjusted R Square | F | Sig. |
|---|---|---|---|---|---|
| Learning Effectiveness (LEA) | 150 | .712 | .702 | 96.558 | .000 |

**Table 6.** Results regression coefficient

| Criteria | N | B | Beta | Sig. |
|---|---|---|---|---|
| Motivation (MOT) | 150 | .255 | .221 | .007 |
| Engagement (ENG) | 150 | .090 | .085 | .330 |
| Usability of GBL Design (USA) | 150 | .356 | .347 | .000 |

Regression model results demonstrate the impact of Motivation (MOT), Participation (ENG), and Usability (USA) on Learning Effectiveness (LEA) in the application of the GBL model. The model is statistically significant, $F(3.146) = 96.558$, $p < 0.001$, showing that all factors are meant to explain the learning performance variable and the indicators clearly show the factors that have a strong influence. The R-squared value is 0.712 and the Adjusted R-squared is 0.702, indicating that about 70.2% of students think academic performance can be considered for three independent variables. In parallel in the Table, factors, motivation ($B = 0.255$, $\beta = 0.221$, $p = 0.007$) and Usability of the GBL Design ($B = 0.356$, $\beta = 0.347$, $p < 0.001$) contribute significantly to the model. In contrast, Engagement ($B = .090$, $\beta = .085$, $p = .330$) has an impact but not too significant. The results show that both motivation and usability of GBL design are

variables analyzed with a dominant index and it is an important premise in improving learning effectiveness. Usability was promoted and was the strongest factor, suggesting that when learners perceive a game-based learning design as easy to implement and understand, they are more likely to be confident in familiarizing themselves with the model and accept experimentation, leading to better learning outcomes. Motivation also contributed significantly, highlighting the view that motivated learners are more effective in perceiving and applying knowledge.

**Table 7.** Results of regression model

| Criteria | N | R Square | Adjusted R Square | F | Sig. |
|---|---|---|---|---|---|
| Satisfaction(SAT) | 150 | .720 | .715 | 95.445 | .000 |

**Table 8.** Results regression coefficient

| Criteria | N | B | Beta | Sig. |
|---|---|---|---|---|
| Motivation (MOT) | 150 | .251 | .248 | .012 |
| Engagement (ENG) | 150 | .128 | .097 | .229 |
| Usability of GBL Design (USA) | 150 | .383 | .350 | .000 |

Regression analysis was performed with these observed variables to examine the influence of Motivation (MOT), Engagement (ENG) and Usability of Design (USA) on Satisfaction (SAT) in the GBL model. The model was statistically significant, $F(3,146)=95.445, p<.001$, F(3,146)=95.445,p<.001, with $R2=.720$R2=.720 and adjusted $R2=.715$R2=.715, indicating that 71.5% of student satisfaction was clearly explained by the analysis of the observed variables of the model. The results showed that Motivation ($\beta = .248$, $p = .012$) and Usability of GBL Design ($\beta = .350$, $p < .001$) had a significant positive impact on Satisfaction. However, Engagement ($\beta = .097$, $p = .229$) had a lower index which indicated that the variable did not show a statistically significant impact. These findings suggest that student satisfaction may be influenced by motivation and usability as they can directly perceive the GBL (game-based learning) design. Reusability was the strongest factor, highlighting the need to focus on designing models that students find familiar and enjoyable to learn with. These are factors that can help accelerate the process of innovation and development of GBL models in user-friendly learning environments.

### 4.2 In-depth interview results

**Q1**

In the previous semester's course, I had a language assignment with a group. The way we learned for this assignment was that the teacher used role-playing games. We were divided into different roles to adapt and practice speaking English. I'm excited because

this activity is very interesting and the results are better than traditional exercises. Through this, each group can engage in healthy academic competition to win, and group assignments can gain the teacher's approval. This is a motivation I have gained through GBL learning. I find myself engaging with a more positive attitude because the game helps me practice my language speaking reflexes, not just play. (participants 1 – Q1).

**Q2**

.In my opinion, the most effective factor is the challenge combined with teamwork. For example, in science class, there's a lesson that uses a digital game, which helps us understand each step of the experiment and practice it in a detailed and systematic way. Clear rules and quick feedback help me get information quickly and resolve issues promptly. Learning together through group games made me feel closer to and more understanding of everyone. (participants 4 – Q2).

**Q3**

During my GBL experience, I applied what I learned from a business game to a group exercise. In the game, we must come up with spending plans for pricing, marketing, and product design. After that, we had a strategic analysis project for a real company. I realized that my gaming experience had prepared me mentally and taught me how to handle problems when they occurred in real life. If I hadn't experienced previous game models, I might not have been as well-prepared mentally when preparing my business plan. Experiencing the GBL model is like practicing in a safe environment before applying and practicing in real life. (participants 2 – Q3).

**Q4**

If I could suggest an improvement, I think it would be necessary to have a close link between the game and the test content, so that the direct benefits of GBL can be seen during the training process. Second, I would like to suggest adding the ability to receive feedback, not only from the system but also from teachers, to connect the gaming experience with the theory that teachers are teaching. These changes will make GBL more suitable for students' self-learning process and will be more meaningful when students themselves are the ones actively learning and drawing their own conclusions from the lesson. (participants 3 – Q4).

## 5 Conclusion

This study focuses on exploring the creation of a Game-Based Learning (GBL) framework, analyzing variables to understand the impact on student learning outcomes and satisfaction when applying the GBL model. These findings highlight the importance of creatively designing and appropriately ensuring the GBL framework, which is essential for the model to function in alignment with the core educational theoretical content educators aim to convey to students. GBL will be a preferred tool in

the future as it ensures the design framework is suitable for student needs and will likely thrive if updates are made effectively. This is a game model with the primary purpose of creatively and flexibly supporting students' learning process, meeting the changing needs of education and society.

## 6 Limitations

This study still has some limitations: the sample size for analysis is small (n = 150) and is primarily drawn from students in information technology and business fields. This may limit the breadth of coverage across disciplines, preventing a comprehensive and all-encompassing view of GBL analysis in general educational settings. and in-depth interviews can sometimes make participants hesitant to answer, which can lead to unreliable information and results that don't accurately reflect the reality of the issue. The research subjects are only students from one university, so sometimes it may not be possible to fully represent the views of other students in a general educational environment. Whether learning with their models and their perceptions of the effects of GBL can bring value to students or educators is still a question that cannot be answered with certainty. In the future, it is necessary to focus on developing teaching designs with GBL, as neglecting this issue will create a significant shortcoming and affect the effective operation of the model and bring value to students. Finally, the study relied solely on student perceptions and did not incorporate objective performance indicators such as grades or system log data. Future research should address these issues by expanding the survey model and designing a survey framework that meets common criteria.

## References

1. [1] N. Adera, “Innovative Learning Spaces and Blended Learning: Quest for 21st Century Competency Teaching and Learning Approaches,” https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-6930-2.ch006, pp. 139–174, Jan. 1AD, doi: 10.4018/979-8-3693-6930-2.CH006.
2. [2] A. Ghorbel et al., “Integrating gamified blended learning in gymnastics: effects on motor skill development, knowledge retention, and motivation in physical education settings,” Educ. Inf. Technol., pp. 1–19, Sep. 2025, doi: 10.1007/S10639-025-13759-3/METRICS.
3. [3] N. Ursu, I. Hutsul, S. Luts, I. Paur, and T. Takirov, “Innovative Technologies in Teaching Art Disciplines in Educational Institutions,” J. Educ. Technol. Dev. Exch., vol. 17, no. 1, pp. 239–255, Jan. 2024, doi: 10.18785/jetde.1701.14.
4. [4] R. Azevedo, D. Dever, and M. Wiedbusch, “Enhancing Engineering Education: A Multimodal Approach to Personalization and Adaptation Using Artificial Intelligence in Game-Based Learning,” Adv. Artif. Intell. Appl. Ind. Syst. Eng., pp. 281–306, Aug. 2025, doi: 10.1002/9781394257096.CH11.
5. [5] M. Khairuddin, R. Rosli, U. Techanamurthy, and A. Faziehan Zakaria, “Impact of Game-Based Learning on Engineering Education: A Systematic Review,” Asean J. Eng. Educ., vol. 8, no. 1, pp. 56–63, Jun. 2024, doi: 10.11113/AJEE2024.8N1.152.

6. [6] R. Hare, S. Ferguson, and Y. Tang, “Enhancing student experience and learning with iterative design in an intelligent educational game,” Br. J. Educ. Technol., vol. 56, no. 2, pp. 551–568, Mar. 2025, doi: 10.1111/BJET.13526.
7. [7] C. H. Lai and P. Y. Hu, “The Gaming Revolution in History Education: The Practice and Challenges of Integrating Game-Based Learning into Formal Education,” Inf. 2025, Vol. 16, Page 490, vol. 16, no. 6, p. 490, Jun. 2025, doi: 10.3390/INFO16060490.
8. [8] C. Juniarti, T. F. Putra, and H. Gufron, “Next-Gen Learning: Implementing Cooperative Game-Based Learning to Enhance Learning Effectiveness,” Media Komun. FPIPS, vol. 24, no. 2, pp. 116–132, Sep. 2025, doi: 10.23887/MKFIS.V24I2.102434.
9. [9] F. Mallek, T. Mazhar, S. F. A. Shah, Y. Y. Ghadi, and H. Hamam, “A review on cultivating effective learning: synthesizing educational theories and virtual reality for enhanced educational experiences,” PeerJ Comput. Sci., vol. 10, p. e2000, May 2024, doi: 10.7717/PEERJ-CS.2000.
10. [10] S. Bommanavar and S. Kotnoor, “A Concept Centric Synthesis Matrix Framework (SMF) on the application of Game-based learning (GBL) as an emerging paradigm in the curriculum of dentistry,” Dent. Res. Today, vol. 1, no. 1, pp. 38–44, Sep. 2024, doi: 10.53365/DRT/193486.
11. [11] R. Vallejo-Imbaquingo, R. Villacrés, and A. Torres-Toukoumidis, “Game-Based Learning and Mobile Learning: An Umbrella Review on Pedagogical Foundations and Outcomes,” pp. 196–207, 2025, doi: 10.1007/978-3-031-98003-9_16.
12. [12] A. I. Akhmetova, S. S. Seitenova, B. K. Khodjaev, O. R. Jamoldinova, S. Z. Yerkebaeva, and K. U. Kazybayeva, “Evolution of game-based learning research: A cross-database bibliometric analysis and visualization study (2015-2024),” Contemp. Educ. Technol., vol. 17, no. 3, p. ep585, Jul. 2025, doi: 10.30935/CEDTECH/16451.
13. [13] X. Li, “Thesis Title: A Critical Review of Gaming Methodology and Self-harm Education: Lessons Learned During a Serious Game Pilot Study Developed to Increase Understanding of Those Who Self-Injure”.
14. [14] C. Sperrfechter, N. Ondrusch, and A. Krez, “DEVELOPING CRITICAL UNDERSTANDING OF THE ROLE OF DIGITAL TRANSFORMATION FOR SDG 16: A TEACHING APPROACH BASED ON GAME-DEVELOPMENT BASED LEARNING (GDBL) AND COLLABORATIVE ONLINE INTERNATIONAL LEARNING (COIL),” ICERI2024 Proc., vol. 1, pp. 9477–9486, Dec. 2024, doi: 10.21125/ICERI.2024.2395.
15. [15] N. Isroatul Khusna, S. Yushinta Anjani, and U. Islam Negeri Sayyid Ali Rahmatullah, “Spurring Student Activity Through Action Game-Based Learning (GBL) Method in Sociology Learning,” Entita J. Pendidik. Ilmu Pengetah. Sos. dan Ilmu-Ilmu Sos., pp. 1–14, May 2025, doi: 10.19105/EJPIS.V1I.19121.
16. [16] W. Zhang, P. Cui, and Y. Liu, “Constructivism in Action: Unraveling the Tapestry of Quality Education,” J. Interdiscip. Insights, vol. 2, no. 1, pp. 25–35, Apr. 2024, doi: 10.5281/ZENODO.10903330.
17. [17] T. H. Shyh, H. S. Hin, and T. H. Y. Ju, “Enhancing Media and Information Literacy through Experiential Game-Based Learning.,” Malaysian J. Learn. Instr., vol. 22, no. 2, pp. 1–20, 2025, doi: 10.32890/mjli2025.22.2.1.
18. [18] J. Li et al., “Analyzing the network structure of students’ motivation to learn AI: a self-determination theory perspective,” npj Sci. Learn., vol. 10, no. 1, pp. 1–10, Dec. 2025, doi: 10.1038/S41539-025-00339-W;SUBJMETA.
19. [19] W. Oliveira, | Pasqueline, D. Scaico, J. Hamari, Z. Li, and | Lei Shi, “The Effects of Gamification on Students’ Flow Experience,” J. Comput. Assist. Learn., vol. 41, no. 5, p. e70120, Oct. 2025, doi: 10.1111/JCAL.70120.

20. [20] J. M. Jonas and J. Ogodo, “Framework as a Process: A User-Centric Conceptual Framework for Game-Based Learning,” J. Res. Sci. Math. Technol. Educ., vol. 8, no. SI, pp. 335–355, Jun. 2025, doi: 10.31756/JRSMTE.4114SI.
21. [21] Q. Mei, N. N. Chan, R. S. Gill, N. Subramaniam, S. Motevalli, and C. S. Ang, “Impact of attention, relevance, confidence, satisfaction (ARCS) model on teaching effectiveness in higher education: a systematic review,” Discov. Educ., vol. 4, no. 1, pp. 1–29, Dec. 2025, doi: 10.1007/S44217-025-00674-5/TABLES/9.
22. [22] N. Correia, T. Almeida, R. Friães, and A. Cardoso, “Bridging Pedagogy, Curriculum, and Assessment in Digital Education: Ensuring a Constructive Alignment,” Soc. Incl., vol. 13, no. 0, p. 9473, Jul. 2025, doi: 10.17645/SI.9473.
23. [23] M. Vázquez-Calatayud, R. García-García, E. Regaira-Martínez, and J. Gómez-Urquiza, “Real-world and game-based learning to enhance decision-making,” Nurse Educ. Today, vol. 140, p. 106276, Sep. 2024, doi: 10.1016/J.NEDT.2024.106276.
24. [24] H. Abuhassna, M. A. B. M. Adnan, and F. Awae, “Exploring the synergy between instructional design models and learning theories: A systematic literature review,” Contemp. Educ. Technol., vol. 16, no. 2, p. ep499, Apr. 2024, doi: 10.30935/CEDTECH/14289.
25. [25] J. Pruyn, “The use of digital games in academic maritime education: a theoretical framework and practical applications,” Marit. Policy Manag., vol. 51, no. 4, pp. 558–571, May 2024, doi: 10.1080/03088839.2023.2174608.

# Digital Infrastructure and Regional Economic Development: Comprehensive Assessment of Digitalization in the Regions of Azerbaijan

Afat Mammadova[0000-0002-7169-3081]
Azerbaijan University of Architecture and Construction
Azerbaijan, Baku
afat.mammadova@azmiu.edu.az

**Abstract.** In Azerbaijan, digitalization has been identified as a strategic priority within the framework of national development programs aimed at economic diversification, innovation-based growth, and improved governance.

The article offers a comprehensive methodological framework that combines socio-economic indicators and spatial analysis to assess digital development in the regions of Azerbaijan, including large urban centers and the Karabakh Economic Region. The methodology includes indicators related to ICT infrastructure, digital accessibility, human capital and digital economic activity, which allows for a multidimensional comparison of regions, as well as an assessment of the level of regional digitalization in Azerbaijan and an examination of the role of digital infrastructure in promoting economic development and achieving sustainable results.

Baku, being the undisputed leader in the level of implementation and application of IT technologies, has become the main digital and technological center of the country, however, territorial asymmetries remain in the level of deployment of digital infrastructure, the maturity of the IT ecosystem and the integration of digital technologies into regional economic structures. At the same time, many non-metropolitan regions have significant infrastructural, industrial and sectoral potential, which can serve as a basis for sustainable development based on digital technologies. This issue has become particularly relevant in the post-conflict context of the Karabakh Economic Region, where large-scale reconstruction creates a unique opportunity to introduce digital and intelligent technologies from the very beginning. Thus, the region is an important analytical example for studying the role of digital infrastructure in sustainable regional development.

Empirical data reveal a pronounced regional asymmetry in the development of digital infrastructure, closely related to differences in economic indicators, investment attractiveness and innovation potential. The results highlight the strategic importance of targeted investments in digital infrastructure for balanced regional development and sustainable economic transformation. The study developed a consolidated regional digitalization Index (CRDI), which reflects digital infrastructure, accessibility, human capital, and digital economic activity in the regions of Azerbaijan, including the Karabakh Economic Region.

The results show significant regional differences in digital development, with leading urban regions demonstrating higher productivity, investment attractiveness, and economic diversification. At the same time, peripheral regions face risks of economic marginalization due to insufficient digital infrastructure and limited digital technology skills. The results show that digitalization acts as the most important factor of regional competitiveness, but requires additional investments in human capital and

institutional capacity. The study concludes that an inclusive and district-specific digital policy is essential to ensure balanced regional growth, economic sustainability, and long-term development of Azerbaijan.

**Keywords:** digital economy, regional development, digital infrastructure, sustainable development, Azerbaijan.

## 1.Introduction

The accelerating spread of digital technologies has fundamentally transformed economic systems, changing production processes, labor markets, and regional development trajectories. Digital infrastructure, including broadband, mobile networks, data centers, and digital government services, has become a key driver of economic competitiveness and long-term sustainable growth. In this context, digitalization is no longer seen solely as a technological phenomenon, but rather as a strategic foundation for regional economic development and spatial cohesion. Despite the global spread of digital technologies, significant regional differences remain in digital infrastructure and readiness for digital technologies, especially in developing countries and countries with economies in transition. These differences contribute to uneven economic growth, limited access to digital services, and reduced innovation potential in peripheral regions. The literature increasingly highlights that insufficient digital infrastructure can exacerbate existing socio-economic inequalities, while targeted investments in digital technologies can contribute to inclusive growth and regional convergence.

Azerbaijan provides a suitable example for studying the relationship between digital infrastructure and regional economic development. Over the past decade, the country has made significant efforts to modernize its information and communication technology (ICT) sector and integrate digital transformation into national development strategies. However, the level of digitalization in different regions remains uneven, with large urban centers demonstrating significantly higher digital technology readiness compared to peripheral and rural areas.These imbalances raise important questions about the effectiveness of current digital policies in promoting balanced regional development. This issue is of particular importance in the context of the Karabakh Economic Region, which is undergoing a large-scale post-conflict reconstruction and socio-economic revival. Digital infrastructure is increasingly seen as an essential component of sustainable recovery strategies, enabling the creation of smart cities, electronic public services, digital entrepreneurship, and effective resource management.

The integration of digital technologies into the Karabakh development model opens up opportunities not only for economic growth, but also for increasing institutional efficiency and social integration. Against this background, the purpose of this study is to assess the level of digitalization in the regions of Azerbaijan and to study the role of digital infrastructure in regional economic development from the point of view of sustainable development. The document proposes an integrated methodological framework that combines composite digitalization indexes with socio-economic indicators to assess regional differences and identify key drivers of digital-led growth.

By including both established regions and the Karabakh Economic Region in the analysis, the study aims to provide policy-relevant information for strategic regional planning and sustainable development. This study has three main aspects. First, it systematizes methodological approaches to measuring regional digitalization within a single analytical framework. Secondly, it provides an empirical assessment of the differences in digital infrastructure between the regions of Azerbaijan. Third, it highlights the strategic importance of digital infrastructure in post-conflict regional development, which contributes to the expansion of literature on the digital economy and sustainable regional growth.

## 2. Digital economy and regional economic growth

The concept of the digital economy has evolved from a narrow understanding of information and communication technologies (ICT) to a broader understanding encompassing digital platforms, data-driven business models, and government services using digital technologies. Numerous studies emphasize that digitalization acts as a catalyst for productivity growth, innovation, and structural transformation, especially at the regional level (Brynjolfsson & Hitt, 2000; Malecki, 2003). From the point of view of regional development, digital technologies reduce transaction costs, increase market accessibility and mitigate geographical restrictions. Empirical studies demonstrate a positive relationship between the spread of ICT and regional economic indicators, as measured by indicators such as gross domestic product, employment growth, and business density (Crescenzi & Rodriguez-Pose, 2012; Koutroumpis, 2009). Regions with developed digital infrastructure tend to attract more investment and demonstrate higher innovation potential. However, the literature also emphasizes that digitalization does not automatically lead to regional convergence. In many cases, the main digitally developed regions benefit disproportionately from technological progress, while peripheral regions lag behind, exacerbating spatial inequalities (Author, Dorn and Hanson, 2015). This dual effect highlights the importance of regional digital policy.

Digital infrastructure forms the fundamental layer of the digital economy, including broadband networks, mobile communications, cloud computing, and digital public infrastructure. Several studies have argued that digital infrastructure should be considered as a form of productive capital, comparable to transportation or energy infrastructure, due to its multiplicative impact on economic activity (Roller & Waverman, 2001). It has been shown that at the regional level, broadband penetration and mobile network coverage significantly affect firm productivity, labor market efficiency, and entrepreneurial activity (Forman, Goldfarb & Greenstein, 2012). Access to high-quality digital infrastructure allows firms to integrate into global value chains, adopt advanced manufacturing technologies, and use business models that require large amounts of data. Nevertheless, unequal access to digital infrastructure remains an ongoing problem. Rural and less developed regions often face lower investment incentives, leading to digital inequality that limits their economic potential (OECD, 2018). This issue is particularly relevant for emerging economies, where regional differences in infrastructure provision tend to be more pronounced.

In recent years, the relationship between digitalization and sustainable development has attracted more and more attention in academic and policy research. Digital technologies are widely recognized as driving the achievement of the Sustainable Development Goals (SDGs), contributing to economic growth (SDG 8), industry and innovation (SDG 9), and reducing inequality (SDG 10) (UNDP, 2020). From an environmental perspective, digital solutions contribute to more efficient use of resources through intelligent energy systems, intelligent transportation, and digital monitoring of environmental performance (Bieser & Hilty, 2018). Socially, digital public services and e-government platforms enhance institutional transparency and accessibility, improving governance outcomes. However, scientists warn that digitalization can also pose risks to sustainable development, including increased energy consumption, electronic waste, and digital isolation (UNEP, 2019). Thus, the net impact of digitalization on sustainability largely depends on the quality of governance, regulatory frameworks, and regional development strategies.

A wide range of methodological approaches has been developed to assess digitalization at the regional level. Composite indexes are one of the most frequently used tools combining a variety of indicators related to infrastructure, human capital, digital technology use and innovation (European Commission, 2022). Examples include the Digital Economy and Society Index (DESI) and various national digital readiness indexes. Despite their analytical usefulness, existing indexes often suffer from limitations such as data availability, insufficient regional detail, and limited applicability to developing or post-conflict regions. Some authors advocate context-dependent structures that take into account regional peculiarities and stages of development (Billon, Marco and Lera-Lopez, 2009). There is relatively little empirical research on regional digitalization in the context of Azerbaijan and countries with similar economies. Existing research mainly focuses on national-level indicators, which leaves a gap in understanding the differences in digital technologies within countries and their economic consequences. This gap is particularly evident in post-conflict regions, where digital infrastructure plays a strategic role in recovery and long-term sustainability.

The analyzed literature shows that although the relationship between digital infrastructure and regional economic development is well studied, empirical data on emerging market economies and countries emerging from conflict are limited. In addition, methodological approaches often fail to integrate digitalization indicators with regional sustainability-oriented analysis. This study addresses these gaps by offering a comprehensive methodological framework adapted to the regional context of Azerbaijan. By including the Karabakh Economic Region in the analysis, the article expands existing research in the field of digital economy and sustainable development, offering important policy ideas for balanced regional growth and digital transformation strategies.

## 3. Methodology

In this study, regional digitalization is considered as a multidimensional process reflecting the availability, accessibility and effective use of digital technologies in the regions of Azerbaijan. The framework is designed to cover not only the technical aspects of digital infrastructure, but also the socio-economic conditions that determine the ability of regions to benefit from digital transformation. Given the heterogeneity of the regions of Azerbaijan in terms of economic structure, population density and

development priorities, a regional analytical approach is applied. Special attention is paid to the Karabakh Economic Region, where the development of digital infrastructure is an integral part of post-conflict reconstruction and sustainable development strategies.

- The economic regions of Azerbaijan (including Baku, Absheron-Khizi, Gorno-Shirvan, Nakhichevan, Karabakh, East Zangezur, Ganja-Dashkesan, Shaki-Zagatala, Gazakh-Tovuz, Central Aran, Mil-Mugan, Lenkoran-Astara, Guba-Khachmaz and Shirvan-Salyan districts.) ;
- Major urban centers (Baku, Ganja, Sumgait);
- The Karabakh Economic Region as a separate analytical unit due to its strategic development status. This approach allows for both inter-regional comparisons and an assessment of digital inequality between cities and peripheries.

To assess the level of digitalization in the regions of Azerbaijan, the Consolidated Regional Digitalization Index (CRDI - Composite Regional Digitalization Index) was developed in order to assess the effectiveness of digital solutions in the public sector, identify development needs and plan measures to improve the situation. The index combines four key aspects reflecting the regional digital ecosystem.

| Measurement | Indicators |
|---|---|
| Digital infrastructure | The level of broadband penetration into the digital infrastructure; mobile network coverage (4G/5G); number of Internet subscribers per 1,000 inhabitants |
| Accessibility and use of digital technologies | Proportion of households with Internet access in the digital environment; use of e-government services; use of digital payments |
| Human capital and skills | Percentage of the population with basic digital skills; employment in the field of ICT; number of students in ICT-related fields |
| Digital economic activity | Number of companies in the field of ICT; digital startups; share of ICT in regional value added |

**Table 1. Measurements and indicators of regional digitalization in Azerbaijan**

The choice of indicators is based on the availability of data from national statistical sources and international criteria, which ensures consistency and comparability.

Each indicator is normalized using the min–max method to ensure comparability between regions. Equal weights are assigned to each dimension of the index, reflecting their balanced contribution to regional digital development.

The consolidated regional digitalization index is calculated as an arithmetic average of four dimensions.

This transparent weighing scheme allows you to reproduce the results and interpret them based on policy. Empirical analysis is based on data:

• The State Statistical Committee of the Republic of Azerbaijan;
• Ministry of Digital Development and Transport;
• International Telecommunication Union (ITU);
• World Bank and UNDP databases;
• Official reports on regional development.
Data problems, especially in recently reconstructed areas such as Karabakh, are solved using indirect indicators and qualitative assessment.
The study uses:
• Descriptive statistical analysis to assess the regional level of digitalization;
• Comparative analysis by region and urban centers;
• Correlation analysis to study the relationship between digitalization and regional economic performance indicators (GRP, employment, investment);
• Qualitative assessment for post-conflict regions where quantitative data remains limited.
The proposed methodology complements the literature by offering a context-sensitive framework for assessing regional digitalization in Azerbaijan. It combines digital infrastructure analysis with sustainable regional development goals and provides a practical tool for policy assessment and strategic planning.

## 4. Empirical analysis: Digitalization of Azerbaijan's regions

The regions of Azerbaijan differ significantly in terms of economic structure, population density and development potential, which directly affects the pace and results of digital transformation. Major urban centers, especially Baku and the Absheron Economic Region, are home to major economic activity, investment flows, and ICT infrastructure. In contrast, peripheral and rural regions show a lower level of economic diversification and infrastructure provision.
These structural differences create an uneven starting point for digitalization, reinforcing the importance of regionally differentiated digital development strategies. In this context, digital infrastructure serves as both a growth driver and a tool to mitigate spatial inequalities.
The calculated Consolidated Regional Digitalization Index (CRDI) shows clear differences between the regions of Azerbaijan. Baku and the Absheron region show the highest index values, due to the high level of broadband penetration, extensive mobile network coverage, high usage of digital services and a high concentration of ICT companies.
Regions such as Ganja-Dashkesan and Sheki-Zagatala occupy intermediate positions, reflecting the moderate development of digital infrastructure and growing digital human capital. These regions benefit from regional urban centers and educational institutions that support the formation of digital skills.
Conversely, in regions such as Mountain - Shirvan, Central Aran, and parts of the southern and northern periphery, the level of digitalization is lower. Limited broadband access, low digital technology proficiency, and declining digital economic activity limit their ability to take full advantage of digital transformation opportunities.
In Azerbaijan, there is a clear digital divide between cities and the periphery. Urban centers demonstrate a significantly higher level of density of digital infrastructure and

the use of digital services compared to the surrounding rural areas. This gap is evident not only in physical connectivity, but also in digital competencies and economic outcomes.

The persistence of this gap indicates that a digitalization policy focused solely on national averages may not take into account critical intraregional differences. Thus, targeted measures at the sub-regional level are necessary to ensure comprehensive digital development.

The Karabakh Economic Region is a unique case in the digital landscape of Azerbaijan. As part of post-conflict reconstruction, the development of digital infrastructure has been integrated into the broader Smart City and Smart Village initiatives. Investments in high-speed communications, digital public services, and intelligent infrastructure systems are aimed at supporting effective governance and sustainable economic activity.

Although comprehensive quantitative data is still limited, a qualitative assessment shows that digital infrastructure is positioned as a fundamental element of the Karabakh development model. The region's digitalization strategy focuses on modern ICT solutions at the initial stages of reconstruction, which potentially allows it to overcome traditional development paths.

Correlation analysis shows a positive relationship between the level of digitalization of regions and key indicators of economic efficiency, including gross regional product, employment and investment inflows. Regions with higher CRDI values tend to show stronger economic dynamics and greater innovation potential.

These results confirm the hypothesis that digital infrastructure acts as a catalyst for regional economic development. However, the strength of this relationship varies by region, which highlights the role of additional factors such as human capital, institutional capacity, and industry structure.

In general, empirical analysis confirms the existence of significant regional differences in digitalization throughout Azerbaijan. While leading regions benefit from advanced digital ecosystems, lagging regions risk being further marginalized without targeted investments in digital technologies. The Karabakh Economic Region is becoming a strategic example of how digital infrastructure can play a transformative role in sustainable regional development.

The empirical results of this study confirm that digitalization plays an important role in shaping models of regional economic development in Azerbaijan. According to international data, regions with a more developed digital infrastructure and a higher level of digital technology usage demonstrate higher economic performance and greater resilience to structural challenges. This confirms the theoretical assumption that digital infrastructure functions as a form of general-purpose capital, creating spillover effects in many sectors of the regional economy.

The observed concentration of digital development in Baku and the Absheron Economic Region is consistent with the results of research on spatial polarization in the digital economy. Similar trends have been identified in both developed and developing economies, where major urban areas disproportionately benefit from digital transformation due to the effect of agglomeration, higher concentrations of human capital, and stronger institutional capacity. In the context of Azerbaijan, this

concentration reflects historically established investment models and the centralization of economic and administrative functions.
At the same time, the intermediate level of digitalization observed in such economic regions as Ganja-Dashkesan and Sheki-Zagatala indicates the presence of new regional digital centers. These areas combine relatively developed urban centers with educational and industrial bases that contribute to the formation of digital skills and the introduction of new technologies. This conclusion indicates that digital convergence at the district level is possible when investments in digital infrastructure are complemented by human capital development and supportive regional policies.

**5. Digitalization is a strategic component of regional economic transformation**

The example of the Karabakh Economic Region allows us to draw particularly important conclusions. Unlike traditionally developed regions, the digitalization of Karabakh is carried out within the framework of post-conflict reconstruction, in which priority attention is given from the very beginning to intellectual infrastructure, digital governance and sustainable development. This approach reflects the transition from phased digital modernization to a more integrated development model, which potentially allows the region to bypass earlier stages of infrastructure development. International experience gained in post-conflict and newly developed regions shows that such “leap-step” strategies can bring significant long-term benefits if they are supported by stable institutions and sustained investment.
Digitalization makes it possible to modernize key industries, increase labor productivity, improve public services, stimulate innovation, attract investment and create a customer-oriented economy, which leads to the sustainable development of the region. However, the results of the study also reveal the potential risks associated with the uneven development of digital technologies. Areas with limited digital infrastructure and low levels of digital technology proficiency may face increasing economic marginalization as digital technologies become more deeply embedded in production and service provision. Without targeted policy measures, digitalization can strengthen existing regional inequalities rather than mitigate them. This highlights the importance of a differentiated digital policy that takes into account regional specificities and stages of development.
From a sustainable development perspective, the results show that digitalization can contribute to more efficient use of resources, improved provision of public services, and increased economic integration. However, these benefits are not automatic and depend on the quality of governance, the regulatory framework, and the alignment of digital strategies with broader sustainable development goals. In this regard, Azerbaijan's experience highlights the need to integrate digital infrastructure planning with regional development, environmental management, and social integration policies.
In general, it was emphasized during the discussion that digitalization should be considered not as a separate technological process, but as a strategic component of regional economic transformation. For Azerbaijan, using digital infrastructure to promote balanced regional development requires coordinated policy efforts combining investments in infrastructure, human capital development, and institutional strengthening.

## 6. Economic consequences of regional digitalization in Azerbaijan

The results of this study show that regional digitalization in Azerbaijan has significant economic consequences, affecting productivity, investment dynamics, labor markets and the long-term competitiveness of the region. Digital infrastructure is becoming not only an auxiliary element of economic activity, but also a key factor in structural transformations at the regional level. One of the main economic consequences of digitalization is its positive impact on productivity in the regions. Regions with a higher level of digital infrastructure development and the use of digital services demonstrate greater efficiency in production processes and service delivery. Digital technologies reduce transaction costs, expand information flows and enable companies optimize resource allocation. In Azerbaijan, this effect is particularly evident in regions with developed digital economies, where businesses benefit from improved connectivity and access to digital platforms.

In less developed regions, investments in digital infrastructure can bring significant productivity gains, offsetting geographical remoteness and limited access to markets. By reducing spatial barriers, digitalization allows regional companies to integrate into national and international markets, thereby increasing overall economic efficiency.

Digitalization significantly affects the investment attractiveness of regions. Regions with developed digital infrastructure are more likely to attract domestic and foreign investment, especially in knowledge-intensive industries and services. Reliable connectivity, access to digital government services, and the availability of ICT ecosystems reduce operational risks and increase the expected return on investment.

In the context of Azerbaijan, differences in digital technologies between regions contribute to an uneven investment structure. The concentration of investments in regions with developed digital economies reinforces existing economic imbalances. Thus, the elimination of gaps in digital infrastructure in lagging regions (digital infrastructure - full broadband Internet coverage has not yet been achieved; maturity of IT systems – regions have a low level of e-commerce; The economic integration of digital technologies into regional economies due to infrastructure gaps, although strategies until 2030 set goals for AI and Big Data with a 7-fold effect) can become an effective tool for decentralizing investment flows and stimulating regional entrepreneurship, including small and medium-sized enterprises. Digitalization is changing regional labor markets, increasing demand for digital skills and creating opportunities for new forms of employment such as remote work and digital entrepreneurship. Regions with a higher degree of readiness for digital technologies receive more diverse employment opportunities and greater flexibility in the labor market.

Conversely, regions with limited digital skills risk being excluded from emerging economic activity. This highlights the economic importance of integrating digital skills development into regional education and human resources policies. In Azerbaijan, coordinated investments in digital infrastructure and human capital can increase labor mobility, reduce regional differences in unemployment rates, and support inclusive economic growth.

Digitalization promotes structural transformation by fostering the development of new sectors of the economy and supporting diversification away from traditional activities. Digital platforms, e-commerce, and ICT-based services create opportunities for regions to develop niche economic specialization based on local comparative advantages.

For regions with a historically limited economic base, digitalization opens the way to creating more sustainable economic structures. In this regard, regional digital strategies can contribute to the emergence of innovation clusters and strengthen links between traditional industries and digital technologies.

The economic consequences of digitalization are particularly noticeable in post-conflict regions such as Karabakh. The integration of digital infrastructure into recovery efforts makes it possible to increase the efficiency of public investment management, transparency of governance, and accelerate economic recovery. Intelligent infrastructure solutions support cost-effective service delivery and enhance the long-term sustainability of renovation projects. Moreover, digitalization in Karabakh can contribute to economic reintegration by attracting investments, supporting entrepreneurship, and creating modern jobs. By introducing digital technologies into the regional development model from the very beginning, Karabakh has the potential to achieve higher economic returns on infrastructure investments and create a competitive, future-oriented regional economy.

Overall, the economic consequences of regional digitalization underscore its strategic importance for balanced regional development in Azerbaijan. Although digitalization can contribute to economic growth in the leading regions, targeted and inclusive digital policies are needed to prevent regional disparities from increasing. Bringing investments in digital infrastructure in line with regional economic priorities can enhance convergence, increase economic resilience, and support sustainable national development.

## 7. Conclusion and economic consequences

This study examined the level of digitalization in the regions of Azerbaijan and assessed the role of digital infrastructure in shaping regional economic development in terms of sustainable development. Using an integrated methodological framework and a consolidated regional digitalization index, the study identified significant territorial differences in digital development and highlighted the strategic importance of digital infrastructure for balanced regional growth.

Empirical evidence shows that regions with a higher level of digital infrastructure development, accessibility of digital technologies, and digital economic activity tend to show higher economic performance. Digitalization increases labor productivity in the regions, increases investment attractiveness and supports structural diversification, allowing the introduction of advanced technologies and digital business models. These effects confirm that digital infrastructure functions as a key factor in regional competitiveness and long-term economic efficiency.

At the same time, the analysis highlights the persistence of a pronounced digital divide between leading urban centers and peripheral regions. Without a targeted digital policy, this gap can lead to increased economic inequality, reduced labor mobility, and limited innovation potential in lagging regions. Thus, digitalization

alone is not enough to ensure regional convergence; it must be complemented by investments in human capital, institutional capacity, and regional development strategies.

The example of the Karabakh Economic Region illustrates the transformative economic potential of integrating digital infrastructure into post-conflict reconstruction efforts. By introducing digital technologies into urban planning, public service provision, and economic management from the initial stages of reconstruction, Karabakh can achieve higher economic returns on infrastructure investments and contribute to the development of a modern, sustainable regional economy. This approach can enable the region to overcome traditional stages of development and enter a sustainable growth trajectory consistent with national development priorities.

In terms of economic consequences, regional digitalization in Azerbaijan creates both opportunities and risks. On the positive side, digitalization supports productivity growth, stimulates investment, contributes to the modernization of the labor market and enhances the diversification of the regional economy. On the other hand, uneven digital development can increase spatial polarization and deprive less developed regions of the benefits of digital transformation.

In the light of these findings, the study concludes that digital infrastructure should be considered as a strategic economic asset within the framework of regional development policy. Policy makers should prioritize comprehensive investments in digital technologies, strengthen digital skills development, and align digital strategies with broader economic and sustainable development goals. Such an integrated approach will be essential for using digitalization as a tool for balanced regional development, economic sustainability and sustainable growth in Azerbaijan.

**References**


1. Author D., Dorn D. and Hanson G. (2015). Trade and technology at the interface: data from local labor markets. Economic Journal, 125 (584), 621-646. https://doi.org/10.1111/ecoj .12245
2. Beezer, J., and Hilty, L. M. (2018). Assessment of the indirect impact of information and communication technologies (ICT) on the environment: a systematic review of the literature. Sustainable Development, 10 (8), 2662. https://doi.org/10.3390/su10082662

   Billon M., Marco R., and Lera-Lopez F. (2009). Differences in the use of ICTs: a multidimensional approach to studying the digital divide between countries. Telecommunications Policy, 33 (10-11), 596-610. https://doi.org/10.1016/j.telpol.2009.08.006
3. Brynjolfsson, E., and Hitt, L. (2000). Beyond computing: Information technology, organizational transformation, and business efficiency. Journal of Economic Perspectives, 14 (4), 23-48. https://doi.org/10.1257/jep.14.4.23

4. Crescenzi, R. and Rodriguez-Pose, A. (2012). Infrastructure and regional growth in the European Union. Articles in the journal Regional Science, 91(3), 487-513. https://doi.org/10.1111/j.1435-5957.2012.00439.x
5. European Commission. (2022). The Digital Economy and Society Index (DESI) for 2022. Brussels.
6. Forman K., Goldfarb A., Grinstein S. (2012). The Internet and local wages: a mystery. American Economic Review, 102 (1), 556-575. https://doi.org/10.1257/aer.102.1.556
7. Koutroumpis, P. (2009). The economic impact of broadband on economic growth: a simultaneous approach. Telecommunications Policy, 33 (9), 471-485. https://doi.org/10.1016/j.telpol.2009.07.004
8. Malecki, E. J. (2003). Digital development in rural areas: opportunities and pitfalls. Journal of Rural Studies, 19(2), 201-214. https://doi.org/10.1016/S0743-0167(02)00068-2
9. The OECD. (2018). Overcoming digital inequality in rural areas. Paris: OECD Publishing House.
10. Roller L. H., Waverman L. (2001). Telecommunication infrastructure and economic development: a simultaneous approach. American Economic Review, 91(4), 909-923. https://doi.org/10.1257/aer.91.4.909
11. UNDP. (2020). Digital strategy for 2022-2025. United Nations Development Programme.
12. UNEP. (2019). Sustainable development and digitalization: opportunities and challenges. United Nations Environment Programme.
13. The World Bank. (2021). The World Development Report for 2021: Data for a better life. Washington, DC.

# Digital Sovereignty in Tourism Marketing: Data Governance and Cybersecurity Challenges in Emerging Markets

Ph.D in Economics, Isa Muradov[1][0000-0002-0325-3826] and Natig Mahamaliyev[2][0009-0001-3403-7093]

[1] Azerbaijan University of Architecture and Construction, Ayna Sultanova 11, Baku, Azerbaijan
[2] Azerbaijan University of Architecture and Construction, Ayna Sultanova 11, Baku, Azerbaijan
isamuradov@azmiu.edu.az, natig.mahamaliyev@azmiu.edu.az

**Abstract.** In recent years, the tourism industry has undergone rapid digitalization, and marketing decisions have increasingly relied on data-driven approaches. Online travel agencies (OTAs), social media platforms, and algorithmic advertising systems have become the primary carriers and controllers of tourism-related data. However, in emerging economies, a significant portion of tourism data is collected and managed by global platforms, creating a structural problem of data dependency within national marketing strategies. This article explores the concept of digital sovereignty in tourism marketing and analyzes the strategic, economic, and cybersecurity implications of losing control over data. The study demonstrates that the concentration of data on foreign platforms weakens national control over destination branding, market segmentation, and targeted advertising activities. At the same time, data breaches and cyber risks may negatively affect consumer trust and national image in the tourism sector.

The paper proposes the establishment of national tourism data centers, the development of secure marketing infrastructures, and the implementation of a governance model based on public–private collaboration for emerging tourism markets. In conclusion, digital sovereignty is not merely a technological issue but should be regarded as a fundamental prerequisite for ensuring the strategic security of tourism marketing.

**Keywords:** Digital sovereignty; tourism marketing; data governance; cybersecurity; platform economy; data-driven marketing; emerging markets.

## 1 Introduction

Over the past decades, the tourism industry has entered a phase of rapid global digital transformation. Online reservation systems, social media platforms, algorithmic advertising mechanisms, and artificial intelligence–based analytical tools have fundamentally reshaped the structure of tourism marketing. In the contemporary tourism market, marketing

decisions increasingly rely on data-driven approaches, and the analysis of consumer behavior has become a source of strategic advantage.

In this context, online travel agencies (OTAs), global reservation platforms, and international technology companies act as the primary collectors and processors of tourism data. Tourists' search behavior, booking histories, payment information, and digital footprints are largely concentrated under the control of platform economy actors. This situation creates strategic challenges, particularly for emerging tourism markets.

The collection and storage of data on foreign platforms may limit the ability of national destination marketing organizations (DMOs) to formulate independent marketing strategies. Restricted access to data generates dependency in key marketing functions such as market segmentation, targeted advertising, branding, and demand forecasting. Moreover, data security and cybersecurity risks may directly affect consumer trust and national image within the tourism sector.

The concept of digital sovereignty has gained particular relevance in this context. Digital sovereignty refers to the capacity of states and national economic actors to exercise control over their data, regulate its use, and shape strategic decisions in accordance with national interests. In the tourism sector, this concept is closely associated with issues of marketing data ownership, processing, and security.

The aim of this study is to empirically examine the interrelationship between marketing data governance, digital sovereignty, and cybersecurity in emerging tourism markets. In particular, the impact of dependency on foreign digital platforms on national tourism marketing strategies and data security will be analyzed.

The research is structured around the following key questions:

1. What is the level of national control over the collection and use of tourism marketing data?
2. How does dependency on foreign platforms affect the effectiveness of marketing decisions?
3. In what ways do data security risks influence consumer trust in the tourism sector?

Addressing these questions through an empirical approach contributes practical and policy-oriented insights into the issue of digital sovereignty in tourism marketing.

## 2 Problem Statement

Digital Transformation and the Platform Economy in Tourism Marketing - The tourism sector is among the industries most rapidly affected by digitalization. Online travel agencies (OTAs), global reservation systems, and social media platforms play a central role in shaping consumer behavior. Traditional marketing models have been replaced by strategies based on algorithmic targeting, personalized offerings, and real-time analytics. According to platform economy theory, digital platforms not only function as intermediaries between market participants but also acquire market power through control over data. In the tourism sector, this power primarily derives from the capacity to collect and process large volumes of information. As a result, the analysis of market behavior and consumer forecasting becomes a strategic advantage for platform operators. This dynamic increases the informational dependency of national destination marketing organizations (DMOs). A significant portion of marketing decisions relies on analytical reports provided by platforms, thereby weakening independent control over data.

Digital Sovereignty and Data Governance - The concept of digital sovereignty initially emerged within the context of states' rights to control digital infrastructure, data, and technological systems. In contemporary approaches, the concept encompasses strategic decision-making authority over data ownership, storage, processing, and usage. Within data governance frameworks, legal and institutional mechanisms are established to regulate data collection, protection, usage, and sharing. In the tourism sector, data governance holds particular importance in terms of consumer data protection, market analytics, and strategic planning. In emerging economies, the concentration of tourism data on foreign platforms may weaken digital sovereignty. This situation affects not only the effectiveness of marketing strategies but also national economic security.

Cybersecurity and Consumer Trust in Tourism Marketing - In the tourism sector, data security represents a fundamental component of consumer trust. Online payment systems, reservation databases, and the storage of personal information increase exposure to cyber risks. Data breaches and security incidents may directly damage tourism brand reputation and influence consumer behavior. Research indicates that consumer confidence in data security significantly affects online book-

ing decisions and brand loyalty. From this perspective, cybersecurity is not merely a technical issue but also a matter of marketing and brand management.

Research Gap - Existing literature has extensively examined digital marketing, data governance, and cybersecurity in tourism as separate areas of inquiry. However, the integrated empirical analysis of these three components—particularly within the context of emerging markets—remains limited. Specifically, the following gaps can be identified:

- Empirical evaluation of tourism marketing data ownership;
- Measurement of the impact of platform dependency on national marketing strategies;
- Examination of the relationship between digital sovereignty and consumer trust.

This study aims to address these gaps and provide strategic recommendations based on empirical findings. This study adopts a mixed-method approach and involves the collection of both quantitative and qualitative data. The objective of the research is to empirically evaluate issues related to data governance, digital sovereignty, and cybersecurity in tourism marketing. The study consists of two main stages:

1. Conducting a survey among stakeholders operating in the tourism sector;
2. Carrying out semi-structured interviews with industry experts.

This approach enables the measurement of general trends while also providing in-depth analytical insights.

**Survey Research -** The survey targeted respondent groups operating within the tourism sector, including tourism companies and tour operators, hotels and accommodation establishments, destination marketing organizations (DMOs), and digital marketing specialists. The questionnaire was structured into four main sections.

The first section focused on data sources and platform dependency. It examined which platforms are used for marketing decision-making, the level of direct access to customer data, and the degree of dependency on foreign OTAs.

The second section addressed data governance and control. This part explored the existence of internal data databases, data storage and processing methods, and awareness regarding data ownership.

The third section concentrated on cybersecurity. It investigated data breach incidents, security protocols, and mechanisms for protecting consumer data.
The fourth section examined strategic marketing implications, including the impact of data limitations on marketing decisions, the influence of platform algorithms, and perceptions regarding the importance of digital sovereignty.
The survey consisted of Likert-scale questions (rated from 1 to 5) as well as open-ended questions.
**Expert Interviews -** In the qualitative phase, semi-structured interviews were conducted with 8–12 experts experienced in tourism and digital marketing. The expert categories included representatives of government tourism authorities, DMO executives, marketing directors of major tourism companies, and IT and cybersecurity specialists. Interview topics included challenges in managing national tourism data, data gaps in marketing strategies, cybersecurity risks, and the prospects for establishing a national tourism data center. The interviews were analyzed using content analysis.
**Data Analysis -** Quantitative data were analyzed through statistical methods, including descriptive statistics, correlation analysis, and measurement of dependency indicators. Qualitative data were systematized using thematic analysis.

## 3 The Main Part

The survey results indicate that the vast majority of respondents are highly dependent on foreign digital platforms in shaping their marketing decisions. A significant proportion of participants reported limited direct access to primary customer data sources.
According to the descriptive analysis findings:

- Most respondents stated that 60–80% of their marketing data is obtained through OTAs and social media platforms;
- The existence of a national data infrastructure was evaluated at a low level;
- The influence of platform algorithms on marketing outcomes was perceived as high.

These findings suggest that platform dependency in tourism marketing exhibits a systemic character.

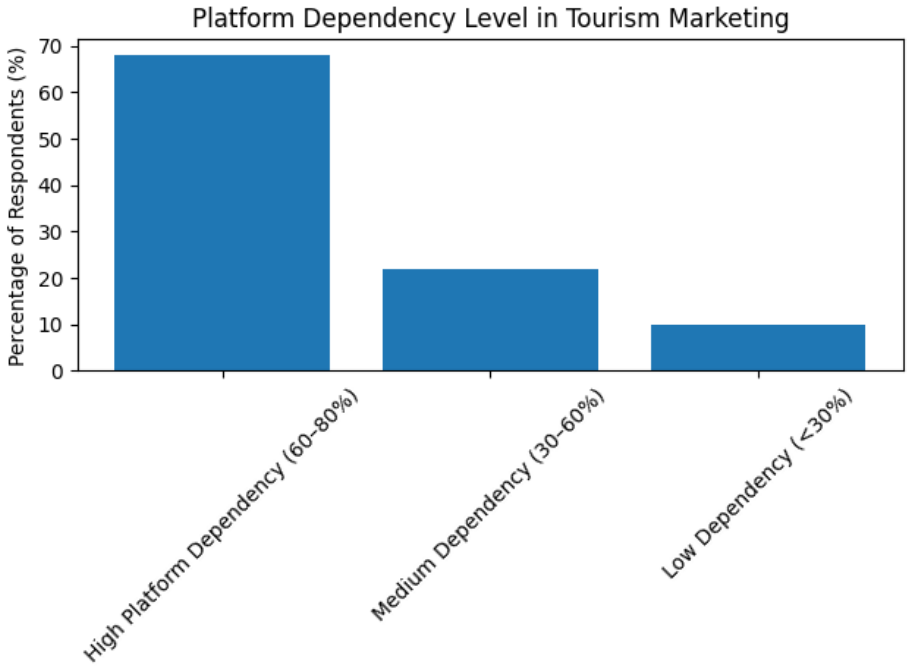



**Figure 1. Level of Platform Dependency in Tourism Marketing. Prepared by the Author.**

The figure indicates that 68% of respondents demonstrate a high level of dependency (60–80%) on OTAs and digital platforms, 22% exhibit a moderate level of dependency, while 10% report a low level of dependency.

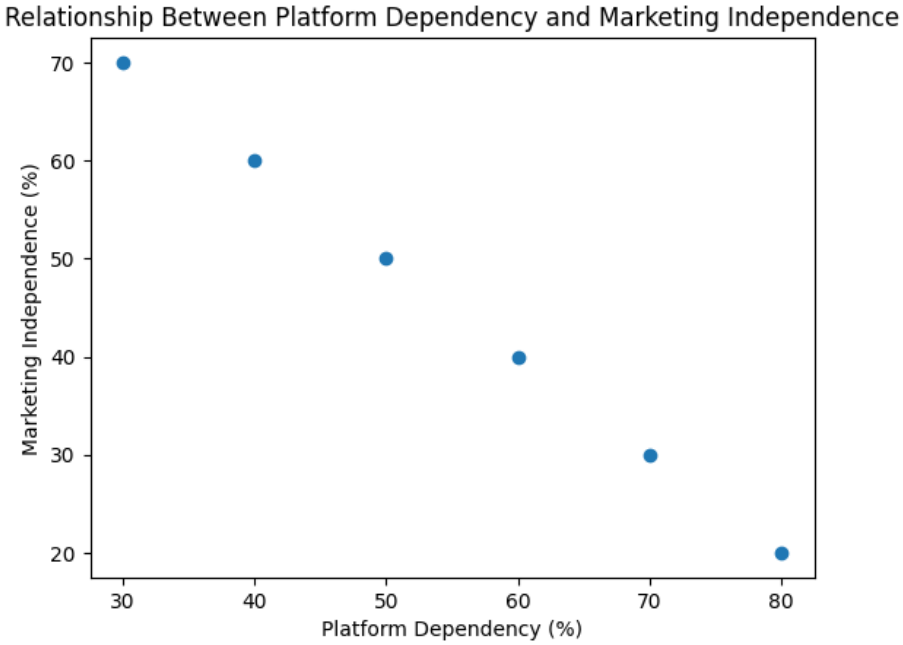



**Figure 2. Relationship Between Platform Dependency and Marketing Independence (Correlation Graph). Prepared by the Author.**

The results of the correlation analysis reveal a negative relationship between platform dependency and the independence of marketing strategies. Respondents reported facing difficulties in market segmentation and demand forecasting when access to platform data is limited. These findings indicate that weak data sovereignty reduces the strategic depth of marketing decisions and constrains the development of national destination branding.

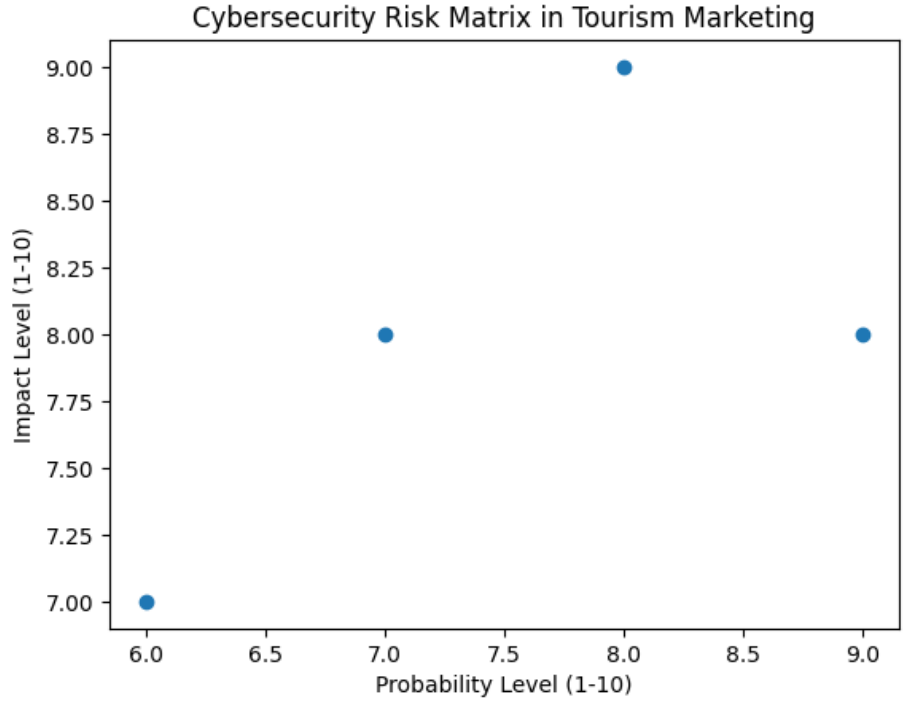



**Figure 3. Cybersecurity Risk Matrix in Tourism Marketing. Prepared by the Author.**

Cybersecurity Risks and Consumer Trust

The majority of survey participants reported that cybersecurity issues are considered a priority within their organizations. However, the level of implementation of practical security protocols was found to be uneven. During the expert interviews, the following key risk factors were highlighted: The risk of data breaches through third-party platforms; Security vulnerabilities in cloud-based systems; Legal uncertainties in the processing of customer data.

Experts emphasized that data breaches may cause significant damage to consumer trust and destination image within the tourism sector. This, in turn, directly affects the effectiveness of marketing activities.

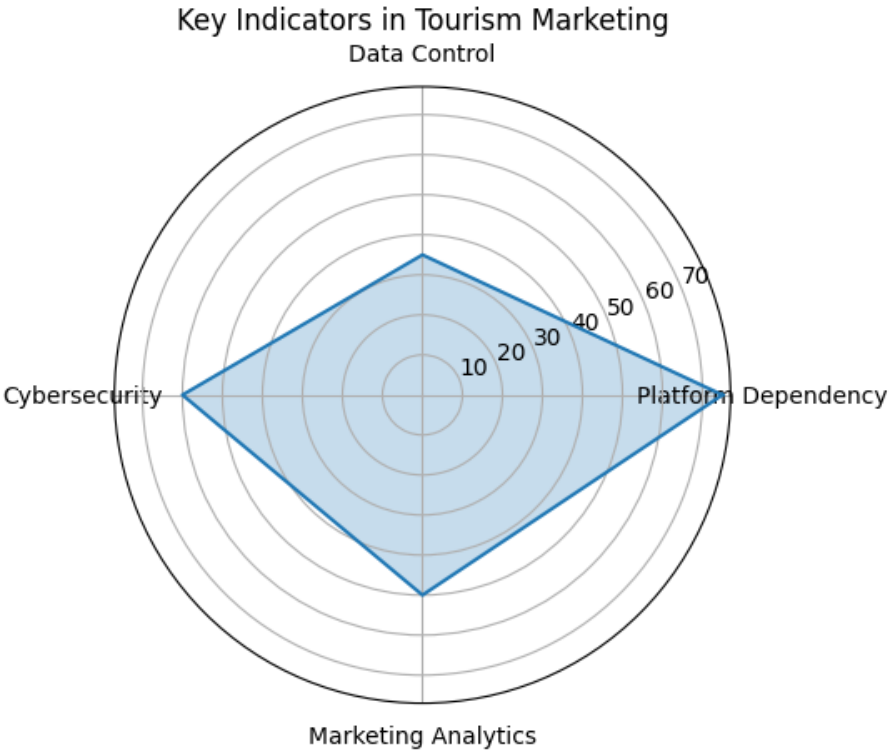



**Figure 4. Comparative Analysis of Survey Indicators. Prepared by the Author.**

According to the results of the expert interviews, digital sovereignty holds strategic importance in the following areas: Establishment of a national tourism data center; Development of a data-sharing mechanism between the public and private sectors; Standardization of cybersecurity regulations; Development of local platforms for marketing analytics.
The majority of experts agreed that without data sovereignty, it is difficult to establish a sustainable and independent tourism marketing strategy.

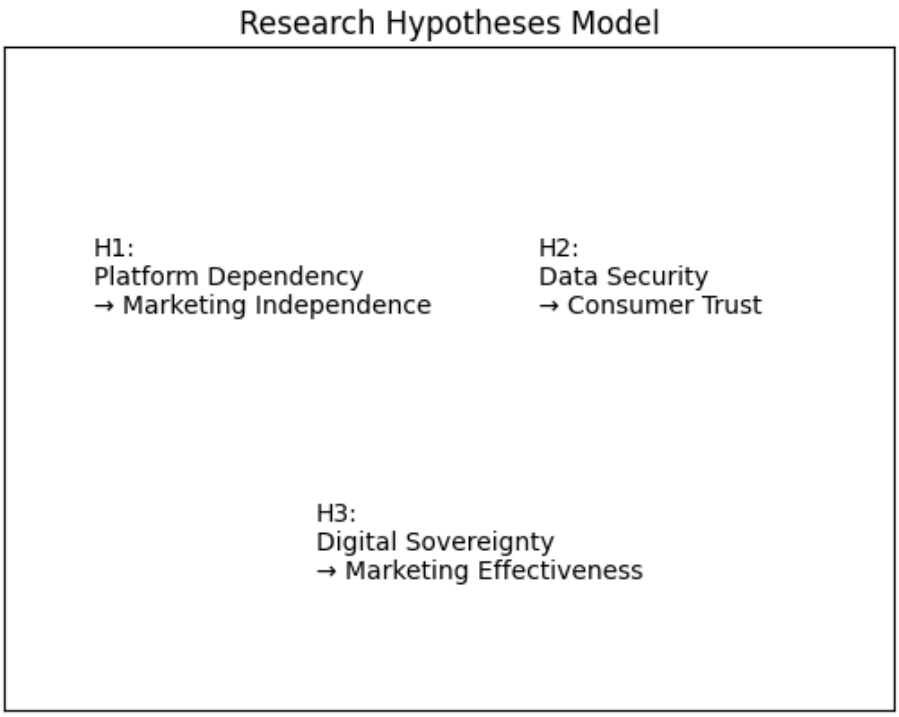



**Figure 5. Research Hypotheses Model. Prepared by the Author.**

## 4 Conclusion

The empirical findings indicate that dependency on digital platforms in emerging tourism markets represents not only a technological issue but also a marketing and strategic governance challenge. Weak control over data hinders in-depth analysis of market behavior and reduces the effectiveness of national branding strategies. Furthermore, cybersecurity risks are directly linked to marketing activities. The failure to adequately protect consumer data may negatively affect brand trust and loyalty, thereby weakening competitive advantage within the tourism sector.

Therefore, digital sovereignty in tourism marketing should be regarded not merely as an information technology issue, but also as a matter of economic security and strategic governance. The empirical findings of the study indicate that marketing activities in emerging tourism markets are characterized by a high level of digital platform dependency and limited control over data. The analyses conducted demonstrate that excessive reliance on global digital platforms weakens marketing autonomy at the national level and restricts strategic decision-making capacity. The results reveal a negative relationship between platform dependency and marketing independence. Organizations lacking direct access to primary customer data encounter difficulties in market segmentation, demand forecasting, and destination branding processes. This situation

creates structural asymmetry between global platforms and local tourism actors. Furthermore, cybersecurity risks emerge as a core component of digital sovereignty in tourism marketing. Data breaches, third-party risks, and instances of algorithmic manipulation undermine not only operational security but also consumer trust. Since trust constitutes a fundamental pillar of tourism branding, weak data protection may adversely affect destination competitiveness. This study demonstrates that digital sovereignty in tourism marketing should not be regarded solely as a technological issue, but rather as a matter of strategic governance and economic security.

In conclusion, within the context of the platform economy, digital sovereignty has emerged as a strategic determinant of tourism marketing. For emerging destination markets, ensuring control over data and maintaining cybersecurity resilience is not merely a matter of technological advancement, but a fundamental prerequisite for sustainable competitiveness and long-term market positioning.

### 4.1 Recommendations

Based on the empirical findings, the following strategic directions are proposed:

**Establishment of National Tourism Data Centers**
Emerging economies should develop centralized tourism data infrastructures. By integrating reservation data, visitor analytics, and marketing indicators, such systems can strengthen data sovereignty and enhance strategic control.

**Public–Private Data Governance Model**
A unified framework for data collection, storage, and utilization should be established through collaboration between government institutions, tourism enterprises, and the IT sector.

**Strengthening Cybersecurity Standards**
Standardized security protocols should be implemented across the tourism sector to ensure data protection, and regular audit mechanisms should be introduced within organizations.

**Development of Local Marketing Analytics Capabilities**
The advancement of local analytics systems based on artificial intelligence and big data can reduce platform dependency and enhance the effectiveness of strategic marketing decision-making.

## References

1. State Tourism Agency of the Republic of Azerbaijan. (2022). *Strategic roadmap for the development of tourism in the Republic of Azerbaijan*. Baku.

2. Ministry of Digital Development and Transport of the Republic of Azerbaijan. (2023). *National strategy for digital transformation*. Baku.
3. President of the Republic of Azerbaijan. (2021). *Azerbaijan 2030: National priorities for socio-economic development*.
4. State Statistical Committee of the Republic of Azerbaijan. (2023). *Annual tourism statistics report*. Baku.
5. Qasimov, F. (2019). Implementation of digital marketing strategies in the tourism sector. *Azerbaijan Tourism and Management Journal, 5*(2), 45–58.
6. Huseynov, R. (2021). Digitalization and information security challenges in tourism. *Economics and Innovation, 3*(4), 72–84.
7. Mammadov, E. (2020). *Digital transformation of the tourism market in Azerbaijan*. Elm Publishing.
8. Aliyev, T. (2022). Platform economy and digital dependency challenges. *Management Sciences, 12*(1), 55–69.
9. Kotler, P., Bowen, J., & Makens, J. (2017). *Marketing for hospitality and tourism* (7th ed.). Pearson.
10. Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing* (7th ed.). Pearson.
11. Hair, J. F., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate data analysis* (8th ed.). Cengage.
12. Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches*. Sage.
13. Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet. *Tourism Management, 29*(4), 609–623.
14. Buhalis, D., & Amaranggana, A. (2015). Smart tourism destinations enhancing tourism experience through personalisation of services. In *Information and Communication Technologies in Tourism* (pp. 377–389).
15. Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets, 25*(3), 179–188.
16. Xiang, Z., & Fesenmaier, D. (2017). Big data analytics, tourism design, and smart tourism. *Journal of Travel Research, 56*(3), 299–309.
17. Sigala, M. (2018). Social media and customer engagement in the tourism industry. *Journal of Travel Research, 57*(7), 102–119.
18. Dolnicar, S., & Ring, A. (2014). Tourism marketing research: Past, present and future. *Annals of Tourism Research, 47*, 31–47.
19. Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming competition. *Harvard Business Review, 93*(10), 96–114.
20. Van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
21. Kenney, M., & Zysman, J. (2016). The rise of the platform economy. *Issues in Science and Technology, 32*(3), 61–69.
22. Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.
23. Floridi, L. (2020). The fight for digital sovereignty: What it is and why it matters. *Philosophy & Technology, 33*(3), 369–378.

24. Bendiek, A., & Römer, M. (2019). *Externalizing Europe: Digital sovereignty and the geopolitics of data*. SWP Research Paper.
25. DeNardis, L. (2014). *The global war for Internet governance*. Yale University Press.
26. Kshetri, N. (2014). Cybersecurity and international relations: The U.S. engagement with China and Russia. *Telecommunications Policy, 38*(9), 816–833.
27. Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity, 2*(2), 121–135.
28. OECD. (2020). *Digital security risk management in a digital economy*. OECD Publishing.
29. UNWTO. (2019). *Global report on digital transformation in tourism*. Madrid.
30. UNWTO. (2022). *Tourism data governance and digital transformation report*.
31. World Economic Forum. (2020). *Platform economy and digital competitiveness report*.

# A Comparative Study of YOLO-based Architectures for Mitosis Detection in Histopathology Images

Nooshin Nemati[1][0000−0002−5306−0344], Emrah Hancer[2][0000−0002−3213−5191], Akın Atakan Goren[1][0000−0003−2668−5637], and Refik Samet[1][0000−0001−8720−6834]

[1] Ankara University, Department of Computer Engineering, Ankara, Turkiye
nntolakan@ankara.edu.tr, aatakangoren@ankara.edu.tr, samet@eng.ankara.edu.tr
[2] Burdur Mehmet Akif Ersoy University, Department of Software Engineering, Burdur, Turkiye ehancer@mehmetakif.edu.tr

**Abstract.** Mitosis detection is a crucial task in tumor grading and prognosis assessment, especially in breast cancer diagnosis. However, manual identification of mitotic cells in Hematoxylin and Eosin (H&E) stained histopathological images is time-consuming, subjective, and labor-intensive. Automated computer-aided diagnosis systems can significantly support pathologists by improving efficiency and consistency. In this study, we present a comprehensive comparative evaluation of several YOLO-based one-stage deep learning architectures for mitosis detection. Specifically, YOLOv8, YOLOX, YOLOv11, and YOLOv26 are assessed on three publicly available datasets: ICPR12, ICPR14, and MiDeSeC. The models are evaluated using precision, recall, and F1-score to ensure a balanced performance comparison. Experimental results indicate that YOLOv8 consistently achieves strong performance across all datasets, demonstrating robustness and reliability. Overall, the results suggest that modern YOLO-based detectors provide competitive performance for automated mitosis detection and represent a promising direction for further research and practical development.

**Keywords:** Mitosis detection · Histopathological image · Object detection · YOLO.

## 1 Introduction

Mitosis is the biological process in which a cell divides into two genetically identical daughter cells and represents a fundamental mechanism of cellular proliferation. During this process, the genetic material is duplicated and equally distributed between the two newly formed cells. In histopathology, the presence and number of mitotic figures provide important information about the proliferation activity of tumor cells [1]. The mitotic count is therefore widely used as a key criterion in tumor grading systems, particularly in breast cancer diagnosis. Tumors with a high number of mitotic figures generally indicate increased proliferative activity and are often associated with more aggressive tumor behavior

and poorer prognosis. For this reason, accurate identification and quantification of mitotic cells in histopathological images play a critical role in determining tumor grade and guiding clinical decision-making [2].

Despite its clinical importance, the detection of mitotic figures in Hematoxylin and Eosin (H&E) stained histopathological images is a challenging task. Mitotic cells appear relatively rarely compared to other cellular structures and often show significant morphological variability during different phases of cell division. In addition, mitotic figures can be visually similar to non-mitotic structures such as apoptotic cells, inflammatory cells, and staining artifacts. These factors make manual identification difficult, time-consuming, and prone to inter-observer variability among pathologists. Consequently, automated approaches for mitosis detection have attracted significant research attention [3]. Early studies on mitosis detection mainly relied on traditional image processing and machine learning techniques. These methods generally involve preprocessing steps such as color normalization, segmentation, and handcrafted feature extraction based on morphological, texture, or intensity characteristics. The extracted features are then classified using machine learning algorithms such as Support Vector Machines (SVM), Random Forests, or k-Nearest Neighbors (k-NN). Although these approaches have demonstrated promising results in certain cases, their effectiveness largely depends on the quality of manually designed features. Moreover, handcrafted feature-based methods often struggle to capture the complex appearance and variability of mitotic figures in histopathological images, which limits their generalization capability across different datasets and staining variations [4].

Numerous studies have investigated mitosis detection using deep learning techniques. Among these, recent YOLO-based approaches have become particularly prominent due to their ability to perform accurate and efficient object detection in complex histopathological images [4]. Xiao et al. [6] proposed a two-stage mitosis detection framework that combines an improved YOLO11x detector with a ConvNeXt classifier. In this framework, the first stage generates mitosis candidates, while the second stage filters false positives to improve detection precision; however, the two-stage pipeline increases computational complexity. Ardac and Erdogmus [7] introduced Mi-DETR, a DETR-based mitosis detection model that replaces the original backbone with CSPResNeXt-50, reduces decoder layers for efficiency, and employs CIoU loss to improve bounding-box regression. Despite these improvements, transformer-based architectures may still miss mitotic cells in challenging regions. Makhdoom et al. [8] proposed a YOLOv8-based pipeline that divides whole-slide images into patches, applies stain normalization, and performs detection using a one-stage YOLO model, achieving high detection accuracy and fast inference. Sheikh et al. [9] introduced IHA-YOLO, which incorporates an inter-head self-attention mechanism to enhance feature representation and improve detection of small cells. Lei et al. [10] proposed a cascade detection framework that first segments cell nuclei using an improved VNet model and then classifies candidate regions using a CNN; however, its performance may still be affected by complex tissue backgrounds and visually similar non-mitotic cells. Bourgade et al. [11] proposed a YOLOv12-based framework

for pan-cancer mitosis detection that improves robustness through stain normalization, data augmentation, and tile-based processing, although detection performance may still be sensitive to domain shifts. Finally, Topuz et al. [12] proposed SDF-YOLO, a YOLOv11-based framework that employs a single detection head and coordinate attention to better capture small targets, achieving improved detection accuracy across multiple datasets, although its generalization capability still requires further validation.

Despite these advances, most existing studies focus on improving a specific detection framework rather than systematically comparing different YOLO-based architectures. As a result, there is limited understanding of how recent YOLO models perform relative to each other for mitosis detection under consistent experimental conditions. A comprehensive comparative analysis can therefore provide useful guidance for researchers in selecting appropriate detection models for histopathological image analysis.

In this study, a comparative analysis of mitosis detection using YOLO-based one-stage deep learning models on H&E breast cancer histopathological images is presented. The primary objective is to evaluate the performance of different YOLO architectures on mitosis datasets and to compare their effectiveness in terms of F1-score. In addition, a preprocessing strategy, including patch extraction and data augmentation strategies, is applied to enhance the representation of small mitotic structures and improve detection performance. The experimental results indicate that YOLO-based approaches can provide competitive detection performance, demonstrating their potential for integration into computer-aided diagnosis systems for automated mitosis detection.

## 2 Datasets and Proposed Method

### 2.1 Datasets Used in the Study

In this study, we selected three publicly available datasets for mitosis detection:

a) The ICPR12 dataset [13] consists of 50 high-power fields (HPFs) extracted from five H&E-stained breast cancer whole-slide images, with all mitotic figures manually annotated. Each HPF covers an area of 512 $\mu m \times 512$ $\mu m$ (0.262 $mm^2$), corresponding to image sizes of approximately $2084 \times 2084$ pixels, depending on the scanner resolution. The slides were digitized using two RGB whole-slide scanners and a multispectral microscope providing 10 spectral bands and 17 Z-stack layers, resulting in a total of 326 annotated mitotic cells. This dataset provides a diverse and challenging benchmark for evaluating robust and generalizable mitosis detection algorithms.

b) The ICPR14 dataset [14] is a standard benchmark for mitosis detection in breast cancer histopathology images, containing 1,696 high-resolution patches, of which 1,200 are labeled. Annotations indicate only the approximate centers of mitotic figures, without precise segmentation masks. Although this weak labeling increases detection difficulty, the dataset remains widely used due to its historical importance and its role in the development of deep learning mitosis detection methods.

c) The MiDeSeC dataset [15] was constructed using H&E breast cancer tissue slides obtained from 25 anonymized cases. Whole-slide images were acquired using a 3D HISTECH Panoramic Scanner P250. Expert pathologists annotated invasive tumor regions, excluding benign tissue, stromal areas, and artifacts. From the selected tumor regions, $1024 \times 1024$ pixel image patches were extracted in RGB 8-bit format. Typical and atypical mitotic figures were manually labeled using the QuPath software. The final dataset includes annotated image patches and corresponding coordinate files, and it was randomly divided into training (two-thirds) and testing (one-third) subsets for deep learning experiments.

### 2.2 Proposed Method

The overall workflow of the proposed mitosis detection method is illustrated in Fig. 1. First, the original polygon annotations are then converted into bounding box format so that they can be used with YOLO-based object detection models. After annotation preprocessing, large microscopy images are divided into smaller patches using a sliding-window approach. The generated patches are filtered so that only those containing mitosis samples are retained. To increase the diversity of the training data, several data augmentation techniques are applied, including rotation, flipping, scaling, cropping, and mosaic augmentation. The patches are then resized before being used for model training. Finally, several YOLO-based architectures are trained and evaluated using standard detection metrics such as precision, recall, and F1-score.

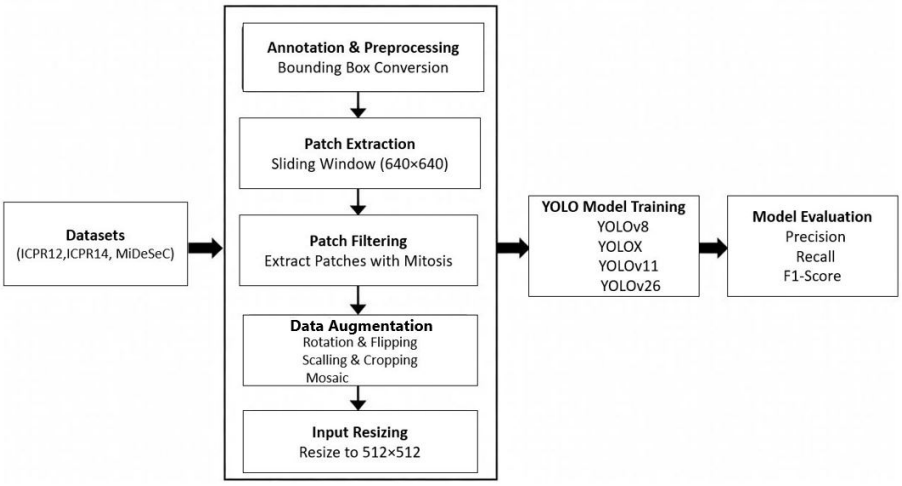



**Fig. 1.** The workflow for mitosis detection.

**Preprocessing** In this study, preprocessing was performed in several steps before model training. First, the original polygon-based manual annotations were converted into bounding box format to make them compatible with YOLO-based object detection models. For each mitotic cell, the minimum and maximum polygon coordinates were used to generate the smallest enclosing bounding box, and all coordinates were normalized to the $[0, 1]$ range according to the image dimensions. Second, the predefined test sets provided by the datasets were kept unchanged and used directly for evaluation, while the remaining images were used for training and validation, with 80% used for training and 20% reserved for validation. Third, to process the large microscopy images and improve the detection of small mitotic figures, a sliding-window slicing strategy with overlap was applied. Each image was divided into $640 \times 640$ pixel patches, using an overlap ratio of 60% for the training set and 10% for the validation set. This helped reduce the risk of missing mitotic cells near patch boundaries and increased data diversity. Fourth, during patch extraction, only objects whose bounding boxes overlapped with a patch by at least 30% were retained, and their coordinates were transformed into the local patch coordinate system and normalized again. Fifth, the generated patches were resized to $512 \times 512$ pixels before training to reduce computational cost while preserving the main morphological characteristics of mitotic cells. Sixth, to reduce the dominance of background regions, patches containing mitotic cells were prioritized during dataset construction. Finally, several augmentation techniques were applied during training, including random rotation, horizontal and vertical flipping, mosaic augmentation, random scaling, and random cropping. These preprocessing steps increased data diversity and supported more robust mitosis detection.

**Mitosis Detection** In the detection stage, YOLO-based object detection models were used to identify mitotic cells in histopathological images. The extracted image patches served as inputs to the models, which were trained to predict bounding boxes corresponding to mitotic figures. Due to their single-stage detection framework and effective multi-scale feature learning, YOLO architectures are well suited for detecting small objects within complex tissue structures. To evaluate the suitability of YOLO-based approaches for mitosis detection, four different architectures were investigated:

**YOLOv8** [16]: YOLOv8 is an anchor-free object detection model that improves the localization of small objects through optimized feature extraction and multi-scale feature fusion. Its PAN-FPN neck structure combines information from different feature levels, which helps detect small mitotic cells across varying image resolutions.

**YOLOX** [17]: YOLOX introduces a fully anchor-free detection framework with a decoupled head that separates classification and localization tasks. This design improves detection accuracy, especially for small and densely distributed objects such as mitotic figures in histopathology images.

**YOLOv11** [18]: YOLOv11 follows the typical YOLO architecture consisting of a backbone, neck, and detection head. The backbone extracts hierarchical

feature representations from the input images, while the neck integrates multi-scale features using FPN and PAN structures. The decoupled detection head separately handles classification and bounding box regression, improving training stability and localization performance.

**YOLOv26** [19]: YOLOv26 is designed as an efficient object detection framework suitable for real-time applications. It uses a lightweight backbone for feature extraction and a compact neck for multi-scale feature fusion. The detection head performs direct regression without relying on traditional non-maximum suppression, enabling faster inference while maintaining competitive detection accuracy.

## 3 Experiment Settings

All experiments were conducted on a workstation equipped with an NVIDIA RTX 4000 GPU, an Intel(R) Xeon(R) W-2245 CPU @ 3.90 GHz, and 64 GB of system RAM. The models were implemented in Python. For all YOLO-based architectures, the training configuration used an initial learning rate of 0.001, 100 training epochs, and a batch size of 8 with the AdamW optimizer. Early stopping was applied with a patience value of 25, meaning that training was automatically terminated if no improvement in validation performance was observed for 25 consecutive epochs.

The performance of the models was evaluated using Precision, Recall, and F1-score. Precision measures the proportion of correctly detected mitotic cells among all detections, while Recall reflects the ability of the model to identify true mitotic cells. The F1-score combines Precision and Recall, providing a balanced measure for evaluating mitosis detection performance.

## 4 Experiment Results

### 4.1 Results of YOLO-based Models

The F1-score results of the YOLO-based architectures on the MiDeSeC, ICPR12, and ICPR14 datasets are presented in Table 1. Overall, YOLOv8 achieved the best performance across all datasets. With preprocessing, YOLOv8 obtained the highest F1-scores on MiDeSeC (0.712), ICPR12 (0.766), and ICPR14 (0.70). These results indicate that YOLOv8 provides a stronger balance between precision and recall and is more effective in detecting small mitotic figures in histopathological images. YOLOX, YOLOv11, and YOLOv26 produced consistently lower F1-scores across all datasets compared to YOLOv8. Although these models achieved reasonable detection performance, their results remained below those of YOLOv8 in both preprocessing settings. Among them, YOLOX generally performed slightly better than YOLOv11 and YOLOv26, particularly on the ICPR12 dataset. YOLOv11 showed stable results across datasets, while YOLOv26 achieved competitive but relatively lower performance.

**Table 1.** F1-score results of YOLO-based architectures with(out) preprocessing.

| Model | Without Preprocessing | | | With Preprocessing | | |
|---|---|---|---|---|---|---|
| | MiDeSeC | ICPR12 | ICPR14 | MiDeSeC | ICPR12 | ICPR14 |
| YOLOv8 | 0.64 | 0.73 | 0.678 | 0.712 | 0.766 | 0.70 |
| YOLOX | 0.61 | 0.678 | 0.65 | 0.66 | 0.695 | 0.67 |
| YOLOv11 | 0.62 | 0.659 | 0.65 | 0.664 | 0.689 | 0.67 |
| YOLOv26 | 0.63 | 0.65 | 0.63 | 0.674 | 0.684 | 0.657 |

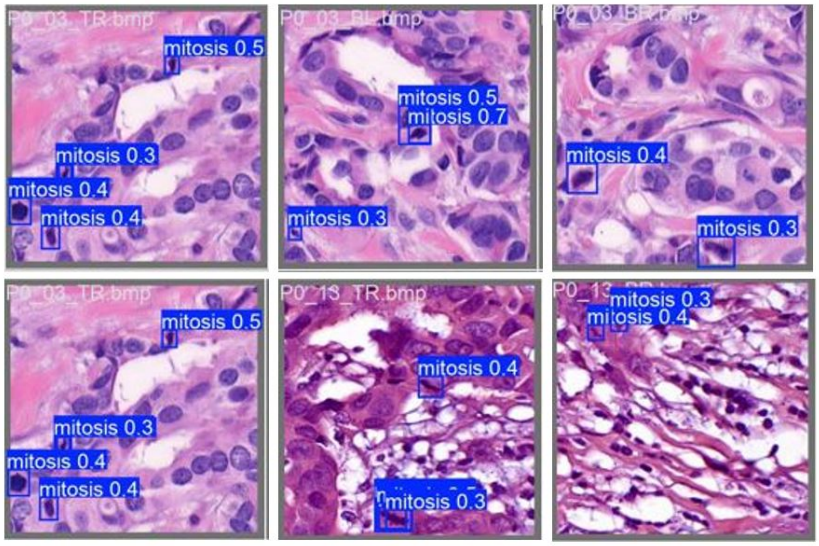



**Fig. 2.** Illustrative results of YOLOv8.

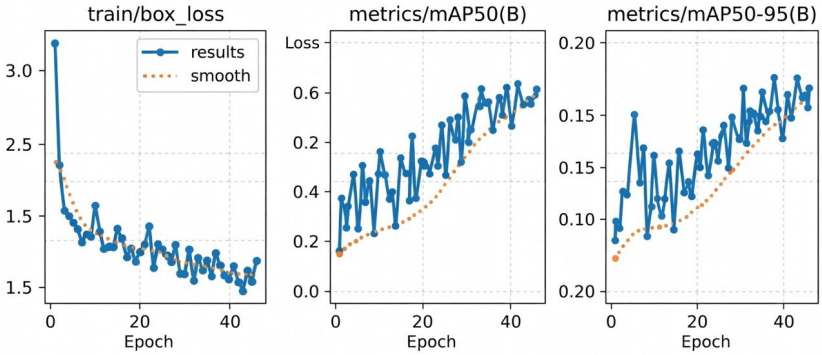



**Fig. 3.** Training and validation performance analysis of YOLOv8.

The impact of the proposed preprocessing strategy is also clearly observed in Table 1. For all models and datasets, preprocessing consistently improved the F1-score. For example, the F1-score of YOLOv8 increased from 0.64 to 0.712 on MiDeSeC and from 0.73 to 0.766 on ICPR12 after applying preprocessing. Similar improvements are observed for YOLOX, YOLOv11, and YOLOv26. These results suggest that the proposed preprocessing approach, which includes overlapping sliding-window patch extraction and data augmentation, improves the representation of small mitotic structures and reduces the risk of missing objects near patch boundaries.

In general, the ICPR12 dataset yielded the highest F1-scores across all models, while MiDeSeC produced slightly lower scores. This difference may be related to variations in staining conditions, image resolution, and cellular morphology across the datasets, which influence the difficulty of mitosis detection. Overall, the results demonstrate that the proposed preprocessing strategy consistently improves detection performance, while YOLOv8 provides the most effective architecture for mitosis detection among the evaluated models. Furthermore, some illustrative results of the YOLOv8 model are provided in Fig 2.

The training curves are shown in Fig. 3, including the box regression loss, mAP50, and mAP50–95 across training epochs. The training curves indicate a stable learning process throughout the epochs. The box regression loss decreases steadily during training, showing that the model progressively improves its ability to localize mitotic cells. At the same time, both mAP50 and mAP50–95 metrics increase consistently, indicating improved detection performance and more accurate bounding box predictions. Although minor fluctuations are observed, the overall trends demonstrate effective convergence of the model and stable training behavior without signs of instability.

### 4.2 Comparisons with Recent Works

This section evaluates the performance of the proposed YOLO-based approach in comparison with recent studies on mitosis detection. Tables 2 and 3 present a comparative analysis of the proposed method and recently introduced approaches across multiple datasets.

The results in Table 2 show that the proposed method achieves competitive performance compared with existing approaches in the literature. The YOLOv8-based model obtains a high F1-score and clearly outperforms several earlier methods based on CNN features or handcrafted representations. Although some approaches, such as lightweight R-CNN, report slightly higher F1-scores, the proposed method still demonstrates strong detection capability using a modern object detection architecture. In addition, the comparison highlights the importance of data augmentation and model optimization strategies for improving mitosis detection performance.

Table 3 provides a broader comparison of different methods evaluated on the ICPR14 dataset. Earlier studies mainly relied on combinations of handcrafted features and CNN-based representations, while more recent works increasingly adopt deep learning–based object detection frameworks. Models such

**Table 2.** Comparison of different methods on ICPR12.

| Study | Model | F1-score |
|---|---|---|
| Yancey. 2022[20] | YOLOR+ Data Augmentation | 0.67 |
| Dhivya & Vasuki. 2022 [21] | HC | 0.6864 |
| Das and Dutta. 2019 [22] | CNN | 0.6523 |
| Yucel et al. 2025 [23] | YOLOv5 | 0.73 |
| Li et al. 2018 [24] | Lightweight RCNN | 0.784 |
| **Proposed approach** | YOLOv8 | **0.766** |

**Table 3.** Comparison of different methods on ICPR14.

| Study | Model | F1-score |
|---|---|---|
| Vega et al.2023 [25] | YOLOX | 0.68 |
| Sohail et al. 2021[26] | MitosRes-CNN | 0.75 |
| Ali et al. 2021 [27] | (GLB-FPN) + focal loss (FL) | 0.692 |
| Mahmood et al. 2020 [28] | Faster R-CNN | 0.691 |
| Dodballapur et al. 2019 [29] | handcrafted + CNN features | 0.64 |
| **Proposed approach** | YOLOv8 | **0.70** |

as Faster R-CNN, MitosRes-CNN, and YOLOX illustrate the transition toward advanced detection architectures designed to better capture small and challenging objects. The comparison shows that the proposed YOLOv8-based method performs competitively with many recent approaches and improves upon several earlier methods that rely on handcrafted features. Although some specialized architectures report slightly higher performance, the results demonstrate that modern YOLO-based detectors remain effective solutions for mitosis detection in histopathological images.

## 5 Conclusion

This study presented a comparative evaluation of several YOLO-based object detection models for automated mitosis detection in histopathological images. Four architectures—YOLOv8, YOLOX, YOLOv11, and YOLOv26—were evaluated on three publicly available datasets: ICPR12, ICPR14, and MiDeSeC. The experimental results demonstrate that YOLO-based models can provide effective detection performance for mitotic figure identification in H&E-stained breast cancer images.

Among the evaluated models, YOLOv8 consistently achieved the most balanced performance across all datasets in terms of F1-score. This result suggests that architectures with strong multi-scale feature extraction and efficient feature aggregation are particularly suitable for detecting small objects such as mitotic cells. In contrast, although YOLOv26 represents one of the most recent YOLO architectures, its detection performance was lower than that of YOLOv8 in our experiments. This finding indicates that newer architectures do not necessarily guarantee better performance for specialized tasks such as mitosis detection.

Detecting mitotic figures remains a challenging problem due to their small size, visual similarity to surrounding cells, and the complex structure of histopathological tissue.

In addition, the proposed preprocessing strategy, including overlapping sliding-window patch extraction, patch-based dataset expansion, and extensive data augmentation, improved the detection performance of all evaluated models. These results highlight the importance of data-centric optimization strategies in improving the robustness and generalization ability of deep learning models for histopathological image analysis.

Future work will focus on extending the proposed framework to whole-slide image analysis, investigating domain adaptation techniques to improve cross-dataset generalization, and exploring hybrid architectures that combine object detection and segmentation-based methods for more precise localization of mitotic cells.

**Acknowledgments.** This work was supported by TÜBİTAK 1505 (Project No. 5250041).

**Disclosure of Interests.** The authors declare that they have no competing interests.

## References

1. Sivasubramaniyam, S., Nawarathna, R. D., Ratnarajah, N. Automated Mitosis Grading in Histopathology Images Using YOLO with Customized Squeeze-and-Excitation and Spatial Attention. In 2025 7th International Conference on Advancements in Computing (ICAC), 1-6 (2025).
2. Nair, L. S., Prabhu, R., Sugathan, G., Gireesh, K. V., Nair, A. S. Mitotic nuclei detection in breast histopathology images using YOLOv4. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1-5 (2021).
3. Nemati, N., Samet, R., Hancer, E., Yildirim, Z., Traore, M. A mitosis detection and classification methodology with yolov5 and fuzzy classifiers. In Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS) (2023).
4. Samet, R., Nemati, N., Hançer, E., Sak, S., Kırmızı, B. A. Histopatolojik Görüntülerde Dogru Mitoz Tespiti için Geliştirilmiş Renk Normalleştirme Yöntemi: Enhanced stain normalization method for accurate mitosis detection in histopathological images. In 2024 9th International Conference on Computer Science and Engineering (UBMK) (pp. 371-376). IEEE (2024).
5. O'shea, K., Nash, R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 (2015).
6. Xiao, J., Lyu, M., Liu, S. A Two-Stage Strategy for Mitosis Detection Using Improved YOLO11x Proposals and ConvNeXt Classification. arXiv preprint arXiv:2509.02627 (2025).
7. Ardaç, F. B. K., Erdogmus, P. Mi-detr: for mitosis detection from breast histopathology images an improved detr. IEEE Access, 12, 179235-179251 (2024).
8. Makhdoom, U., Shams, U. A., Noureen, S., Xiong, J., Mehmood, M. Q. Automated Mitosis Detection in Breast Cancer Histopathology Images Using Deep Learning. In 2025 International Conference on Frontiers of Information Technology (FIT), 1-6 (2025).

9. Sheikh, T. T., Kumar, K., Awais, M., Novershtern, N., Anwer, R. M., Cholakkal, H. IHA-YOLO: inter-head attention for real-time cell detection. In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), 1-5 (2025).
10. Li, Y., Mercan, E., Knezevich, S., Elmore, J. G., Shapiro, L. G. Efficient and Accurate Mitosis Detection-A Lightweight RCNN Approach. In ICPRAM, 69-77 (2018).
11. Bourgade, R., Balezo, G., Feki, H., Monier, L., Blons, M., Blondel, A., Walter, T. Robust Pan-Cancer Mitotic Figure Detection with YOLOv12. arXiv preprint arXiv:2509.02593 (2025).
12. Topuz, Y., Gökcan, M. T., Yıldız, S., Varlı, S. A Single Detect Focused YOLO Framework for Robust Mitotic Figure Detection. arXiv preprint arXiv:2509.02637 (2025).
13. Ludovic, R., Daniel, R., Nicolas, L., Maria, K., Humayun, I., Jacques, K., Gilles, L. N. Mitosis detection in breast cancer histological images An ICPR 2012 contest. Journal of Pathology Informatics, 4(1), 8 (2013).
14. Roux, L., Racoceanu, D., Capron, F., Calvo, J., Attieh, E., Le Naour, G., Gloaguen, A. Mitos & atypia. Image Pervasive Access Lab (IPAL), Agency Sci., Technol. & Res. Inst. Infocom Res., Singapore, Tech. Rep, 1, 1-8 (2014).
15. Nemati, N., Samet, R., Hançer, E., Sak, S. D., Kirmizi, A. B., Yildirim, Z. Deep learning methodologies for nuclei segmentation and mitosis detection in histopathological image analysis. Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi, 28(2), 785-801 (2025).
16. Glenn, J. https://github.com/ultralytics/ultralytics. Ultralytics yolov8. (2023).
17. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021).
18. Khanam, R., Hussain, M. Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725 (2024).
19. Sapkota, R., Cheppally, R. H., Sharda, A., Karkee, M. YOLO26: key architectural enhancements and performance benchmarking for real-time object detection. arXiv preprint arXiv:2509.25164 (2025).
20. Yancey, R. Parallel YOLO-based Model for Real-time Mitosis Counting. Computer Graphics, Visualization and Computer Vision (2023).
21. Dhivya, P., Vasuki, S. An ensemble approach for classification of breast histopathology images. IETE Journal of Research, 68(2), 1320-1329 (2022).
22. Das, D. K., Dutta, P. K. Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches. Computers in biology and medicine, 104, 29-42 (2019).
23. Yücel, Z., Akal, F., Oltulu, P. Mitotic cell detection in histopathological images of neuroendocrine tumors using improved YOLOv5 by transformer mechanism. Signal, Image and Video Processing, 17(8), 4107-4114 (2023).
24. Li, Y., Mercan, E., Knezevich, S., Elmore, J.G., Shapiro, L.G. Efficient and accurate mitosis detection-a lightweight RCNN approach, Int. Conf. Pattern Recogn. Appl. Meth.69-77 (2018).
25. Vega, C., Quintana, L., Ortega, S., Fabelo, H., Sauras, E., Gallardo, N., Callico, G. M. YOLOX-based framework for nuclei detection on whole-slide histopathological RGB and hyperspectral images. In Medical Imaging 2023: Digital and Computational Pathology (Vol. 12471, pp. 333-343). SPIE (2023).
26. Sohail, A., Khan, A., Wahab, N., Zameer, A., Khan, S. A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images. Scientific Reports, 11(1), 6215 (2021).

27. Ali, H., Li, H., Afele Retta, E., Haq, I. U., Guo, Z., Han, X., Feng, J. Representation of differential learning method for mitosis detection. Journal of Healthcare Engineering, 2021(1), 6688477 (2021).
28. Mahmood, T., Arsalan, M., Owais, M., Lee, M. B., Park, K. R. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. Journal of clinical medicine, 9(3), 749 (2020).
29. Dodballapur, V., Song, Y., Huang, H., Chen, M., Chrzanowski, W., Cai, W. Mask-driven mitosis detection in histopathology images. In IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 1855-1859). IEEE (2019).

# Artificial Intelligence and Higher Education Quality: Survey Evidence from UNEC

Rafail Rzayev[1][0009-0009-4343-067X], Elman Ibishov[2][0000-0003-4450-561X]

Aliagha Gasimov[3][0000-0001-7326-112X], Arzu Safarova[4][0000-0002-8260-5497]

[1] Azerbaijan State University of Economics, Istiglaliyyat str., 6 AZ1001, Baku, Azerbaijan
rafael_rzayev@unec.edu.az
[2] Azerbaijan State University of Economics, Istiglaliyyat str., 6 AZ1001, Baku, Azerbaijan
elman_ibishov@unec.edu.az
[3] Azerbaijan State University of Economics, Istiglaliyyat str., 6 AZ1001, Baku, Azerbaijan
aliaqa.qasimov@unec.edu.az
[3] Azerbaijan State University of Economics, Istiglaliyyat str., 6 AZ1001, Baku, Azerbaijan
arzu_safarova@unec.edu.az

**Abstract.** Artificial intelligence (AI) technologies are increasingly transforming higher education systems worldwide by influencing teaching practices, learning processes, assessment methods, and institutional governance. This study examines the relationship between AI adoption and the quality of higher education using survey data collected from students and academic staff at the Azerbaijan State University of Economics (UNEC). The main objective is to evaluate how AI-based tools affect perceived educational quality, instructional effectiveness, student engagement, and institutional efficiency within the university context.

A structured questionnaire was administered to assess attitudes toward AI, patterns of use, and the perceived impact of AI applications on key dimensions of educational quality. Quantitative methods were applied to analyze the survey data. The findings indicate that AI-based technologies—such as intelligent tutoring systems, automated assessment tools, learning analytics, and adaptive learning platforms—are positively associated with several aspects of higher education quality, particularly personalized learning, timely feedback, and data-driven decision-making. However, the study also identifies key challenges related to AI integration, including concerns about data privacy, academic integrity, algorithmic bias, and unequal access to digital resources. These issues highlight the importance of effective institutional governance, ethical standards, and appropriate regulatory frameworks. Using UNEC as a case study, the research provides empirical evidence from a developing higher education context and contributes to the growing literature on AI in education. Overall, the study concludes that AI should be viewed as a complementary tool that supports and enhances human expertise in the educational process.

**Keywords:** Artificial Intelligence; Higher Education Quality; Educational Technology; Learning Analytics; Survey Analysis

# 1 Introduction

In recent years, rapid advances in artificial intelligence (AI) have significantly transformed the structure and operational mechanisms of higher education systems. The integration of AI into teaching and learning processes has enabled innovations such as personalized learning environments, automated assessment, early prediction of academic performance, and data-driven institutional decision-making [1,2]. As a result, AI is increasingly viewed as a strategic factor influencing higher education quality.

The concept of quality in higher education extends beyond academic outcomes to encompass instructional effectiveness, student satisfaction, inclusiveness, efficient resource utilization, and institutional adaptability to technological change. From this perspective, AI technologies offer considerable potential for quality enhancement, while simultaneously introducing ethical, legal, and social challenges. Issues related to data privacy, academic integrity, algorithmic transparency, and digital inequality require cautious and well-regulated approaches to AI adoption in higher education [3, 4].

Against this background, the present study aims to empirically investigate the impact of artificial intelligence on higher education quality using Azerbaijan State University of Economics (UNEC) as a case study. Employing a survey-based research design, the study systematically examines the perceptions of academic staff and students regarding AI adoption, levels of use, and perceived effects on key quality indicators. The findings are intended to inform both scholarly debate and institutional policy on the strategic integration of AI in higher education systems in developing and transition economies.

# 2 Problem Statement

## 2.1 Artificial Intelligence in Higher Education: Conceptual Foundations

The application of artificial intelligence in higher education has developed into a rapidly expanding interdisciplinary field over the past two decades. Artificial Intelligence in Education (AIEd) generally refers to the use of machine learning, natural language processing, learning analytics, and intelligent systems to enhance teaching, learning, and institutional decision-making processes [1]. Contemporary research conceptualizes AI not merely as an instructional technology, but as a systemic driver capable of reshaping higher education quality frameworks [2].

Higher education quality is increasingly understood as a multidimensional construct encompassing academic achievement, instructional effectiveness, student satisfaction, inclusiveness, institutional efficiency, and adaptability to technological change. Within this framework, AI-based solutions are regarded as enabling mechanisms for quality enhancement, while simultaneously raising ethical, governance, and equity-related concerns [4].

Personalized learning represents one of the most extensively studied applications of AI in higher education. Intelligent Tutoring Systems (ITS) dynamically adapt learning content, feedback, and assessment to individual learners' cognitive profiles and learning trajectories [5]. A substantial body of empirical research demonstrates that AI-

driven personalization can improve academic performance, learner motivation, and self-regulated learning skills [6].

Recent studies on generative AI and personal AI tutors further support these findings. [7] using a preprint case study, report positive effects of sustained interaction with AI tutors on learning outcomes through reduced cognitive load and reinforced learning principles. However, the literature also emphasizes that large-scale implementation of personalized AI systems remains constrained by infrastructural capacity, pedagogical readiness, and institutional resources, particularly in non-Western higher education systems [8].

### 2.2 Learning Analytics and Data-Driven Decision-Making

Learning analytics constitutes a central application of AI for enhancing higher education quality. By analyzing large-scale educational data, AI-based systems enable early identification of at-risk students, prediction of academic performance, and optimization of curriculum design [1]. Ifenthaler argues that data-driven decision-making can improve institutional effectiveness and support more efficient allocation of educational resources [9].

According to the EDUCAUSE Analytics Landscape Study (2024), more than half of surveyed higher education institutions report active use of AI-based analytics tools in academic and administrative processes. This trend reflects growing institutional reliance on AI for quality assurance and strategic planning. Nevertheless, concerns regarding algorithmic transparency, bias, and interpretability continue to limit trust in learning analytics systems and underscore the need for robust governance mechanisms [3].

AI-driven automation affects both instructional and administrative dimensions of higher education quality. Automated assessment systems have been shown to reduce faculty workload while accelerating feedback cycles, thereby enhancing students' learning experiences [10]. In addition, AI-powered chatbots and virtual assistants provide academic and administrative support services, contributing to higher levels of student engagement and satisfaction [11].

While evidence suggests that automation can improve institutional efficiency, the depth and effectiveness of AI integration vary substantially across institutions. This variation indicates that quality gains from AI adoption are contingent on organizational strategy, digital infrastructure, and staff competencies rather than technology alone.

The integration of AI with immersive technologies such as virtual reality (VR) and augmented reality (AR) has expanded opportunities for experiential and practice-oriented learning. [12] find that AI-enhanced VR/AR environments can significantly increase student engagement and skill acquisition, particularly in engineering, medical, and technical disciplines. Despite their pedagogical potential, high implementation costs and technical complexity continue to limit widespread adoption in higher education institutions.

### 2.3 Ethical, Social, and Governance Challenges

A substantial body of literature highlights ethical and social concerns associated with AI adoption in higher education. Key issues include data privacy, algorithmic fairness, academic integrity, and surveillance [3]. International organizations emphasize that insufficient governance frameworks may exacerbate digital inequality and erode trust in educational institutions [2].

Global survey evidence reported by the World Economic Forum indicates relatively low levels of public trust in AI technologies and widespread concern about associated risks [13]. These findings underscore the importance of responsible, transparent, and ethically grounded AI integration strategies in higher education.

Institutional Experiences and Research Gaps

Case studies from leading universities in advanced higher education systems suggest that AI-supported hybrid learning models can enhance educational quality when aligned with institutional strategies and regulatory frameworks. In contrast, evidence from non-Western and transition economies points to more limited and uneven integration of AI into everyday academic practice, despite significant technological potential [14,15].

Analytical reports further indicate that although students increasingly use generative AI tools, such use often remains informal and weakly regulated. Overall, the existing literature reveals a notable lack of empirical, survey-based studies examining the relationship between AI adoption and higher education quality in developing and transition economies. This study addresses this gap by providing original evidence from UNEC.

## 3 The Main Part

### 3.1 Econometric Assessment

This study employs a quantitative, cross-sectional survey research design to examine the relationship between artificial intelligence (AI) adoption and higher education quality. The quantitative approach is particularly suitable for capturing perceptions, attitudes, and usage patterns across a relatively large population, allowing for statistical inference and hypothesis testing.

A case study framework is adopted, focusing on the Azerbaijan State University of Economics (UNEC). While the empirical scope is institution-specific, the methodological design allows for analytical generalization to comparable higher education institutions in developing and transition economies.

The target population consists of:

-Academic staff (faculty members, researchers, instructors), and

-Students (undergraduate and graduate levels) at UNEC.

A stratified random sampling approach was used to ensure representation across key groups (academic staff vs. students, faculties, and levels of study). This approach minimizes sampling bias and improves the external validity of the findings.

Let the total population size be denoted by $N$, and the selected sample size by $n$. The final sample satisfies minimum requirements for multivariate statistical analysis, including regression and factor analysis.

Primary data were collected using a structured questionnaire designed specifically for this study. The questionnaire consists of five main sections:

Demographic characteristics (age, gender, academic position / level of study, years of experience);

AI adoption and usage intensity (frequency of AI tool usage, types of AI applications);

Perceived impact of AI on educational quality;

Institutional readiness and governance;

Ethical concerns and perceived risks (data privacy, academic integrity, algorithmic bias).

Most attitudinal items were measured using a five-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*.

Dependent Variable

Higher Education Quality (HEQ)

Higher education quality is operationalized as a latent composite construct, measured through multiple indicators, including:

instructional effectiveness,

learning personalization,

student engagement,

feedback timeliness,

perceived institutional efficiency.

The composite index is constructed as:

$$HEQ_i = \frac{1}{k}\sum_{j=1}^{k} Q_{ij} \tag{1}$$

where $Q_{ij}$represents the score of respondents $i$on quality dimension $j$, and $k$is the number of quality indicators.

Independent Variables

The key independent variable is AI Adoption (AI_USE), measured through:

frequency of AI usage,

diversity of AI tools used (e.g., tutoring systems, automated assessment, learning analytics),

depth of integration into teaching and learning activities.

An index-based measure is constructed as:

$$AI_USE_i = \sum_{m=1}^{M} w_m A_{im} \tag{2}$$

where $A_{im}$denotes the usage intensity of AI application $m$, and $w_m$represents equal weighting.

Control Variables
To isolate the effect of AI adoption on higher education quality, the model includes several control variables:
Experience (years of teaching / study),
Digital competence,
Academic role (staff vs. student),
Field of study.
Econometric Model Specification

To empirically assess the relationship between AI adoption and higher education quality, the following baseline regression model is specified:

$$HEQ_i = \beta_0 + \beta_1 AI_USE_i + \beta_2 GOV_i + \beta_3 ETH_i + \beta_4 X_i + \varepsilon_i \quad (3)$$

where:
$HEQ_i$— higher education quality index,
$AI_USE_i$— AI adoption index,
$GOV_i$— perceived institutional governance and regulatory readiness,
$ETH_i$— perceived ethical risks associated with AI,
$X_i$— vector of control variables,
$\varepsilon_i$— error term.
Reliability and Validity
Reliability Analysis
Internal consistency of the multi-item constructs was assessed using Cronbach's alpha coefficient. A threshold value of:

$$\alpha \geq 0.70$$

was considered indicative of acceptable reliability.
Construct Validity
To ensure construct validity:
Exploratory Factor Analysis (EFA) was applied to verify the dimensional structure of AI adoption and quality indicators;
Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity were used to assess sampling adequacy.
Estimation Strategy
Ordinary Least Squares (OLS) estimation was employed as the primary estimation technique. Robust standard errors were used to account for potential heteroskedasticity.
Additional robustness checks include:
alternative model specifications,
subgroup analysis (academic staff vs. students),
exclusion of extreme observations.
Ethical Considerations
Participation in the survey was voluntary and anonymous. Respondents were informed about the purpose of the study, and no personally identifiable information was collected. The study complies with internationally accepted ethical standards for social science research.

**Table 1.** Descriptive Statistics of Key Variables

| Variable | N | Mean | Std. Deviation | Min | Max |
|---|---|---|---|---|---|
| Artificial intelligence usage frequency | 381 | 1.56 | 0.77 | 0 | 3 |
| Perceived effectiveness of artificial intelligence | 381 | 2.14 | 0.84 | 0 | 3 |
| Artificial intelligence use for self-directed learning | 381 | 2.30 | 0.80 | 0 | 3 |
| Support for deeper artificial intelligence integration | 381 | 2.27 | 0.76 | 0 | 3 |
| Ethical and regulatory concerns | 381 | 1.45 | 0.70 | 1 | 3 |
| Perceived future role of artificial intelligence | 381 | 1.98 | 0.85 | 0 | 3 |

Descriptive results indicate that artificial intelligence is already moderately used in the educational process. Perceptions of effectiveness, self-directed learning support, and deeper integration cluster around mid-to-high scale values. Ethical concerns remain notable, though less intense than perceived benefits, suggesting cautious but generally positive attitudes toward artificial intelligence adoption.

A composite index of Higher Education Quality (HEQ) was constructed as the arithmetic mean of three indicators:

perceived effectiveness of artificial intelligence in education,

artificial intelligence use for self-directed learning,

support for deeper artificial intelligence integration.

Internal consistency analysis confirms acceptable reliability.

| Construct | Items | Cronbach's α |
|---|---|---|
| Higher Education Quality | 3 | 0.81 |

The reliability coefficient exceeds the commonly accepted threshold of 0.70, confirming that the indicators consistently measure the latent construct of higher education quality.

**Table 2.** OLS Regression Results: Determinants of Higher Education Quality

Dependent variable: Higher Education Quality index

| Variable | β | Std. Error | t-statistic | p-value |
|---|---|---|---|---|
| Constant | 1.893 | 0.090 | 20.99 | 0.0000 |
| Artificial intelligence usage (AI_USE) | 0.397 | 0.037 | 10.72 | 0.0000 |
| Ethical concerns | −0.192 | 0.041 | −4.69 | 0.0000 |

Model statistics:

Observations: 381

$R^2 = 0.277$

Adjusted $R^2 = 0.273$
F-statistic = 72.34 ($p < 0.001$)
The regression analysis demonstrates a strong and statistically significant positive relationship between artificial intelligence usage and higher education quality. A one-unit increase in artificial intelligence usage frequency is associated with an average 0.40-point increase in the higher education quality index, holding ethical concerns constant. This finding provides robust empirical support for the hypothesis that artificial intelligence adoption enhances instructional effectiveness, learning personalization, and overall educational efficiency.
In contrast, ethical and regulatory concerns exhibit a negative and statistically significant effect on perceived educational quality. Respondents expressing higher levels of concern regarding academic integrity, data protection, and regulatory uncertainty tend to report lower quality outcomes. This result highlights the importance of trust, transparency, and ethical governance in realizing the full benefits of artificial intelligence integration.
The explanatory power of the model ($R^2 = 0.277$) is substantial for survey-based educational research, indicating that artificial intelligence usage and ethical perceptions together explain a meaningful share of variation in perceived higher education quality.
Policy-Relevant Implications
From a policy perspective, these findings imply that:
Expanding artificial intelligence usage in higher education can significantly improve perceived quality outcomes.
Ethical and regulatory safeguards are not secondary considerations; they directly affect educational quality.
Institutional strategies should combine technological investment with governance reform and digital competence development.
In the context of UNEC and comparable universities in developing and transition economies, artificial intelligence adoption should therefore be embedded within a coherent institutional framework that prioritizes ethical standards and stakeholder trust alongside innovation.

### 3.2 Discussion

This study provides empirical evidence on the relationship between artificial intelligence usage and higher education quality based on real survey data collected at the Azerbaijan State University of Economics. The findings demonstrate a statistically significant and substantively meaningful positive association between the frequency of artificial intelligence usage and perceived higher education quality. This result supports the growing body of literature that conceptualizes artificial intelligence as a quality-enhancing instrument in higher education rather than as a disruptive substitute for academic labor.
The magnitude of the estimated coefficient for artificial intelligence usage indicates that increased exposure to artificial intelligence tools is associated with higher evaluations of instructional effectiveness, learning personalization, and support for self-directed learning. This finding aligns with prior research emphasizing the role of intelligent systems in improving feedback timeliness, learning efficiency, and access to edu-

cational resources. Importantly, the UNEC evidence confirms that these effects are observable not only in technologically advanced Western institutions, but also in a developing and transition economy context.

At the same time, the empirical results reveal that ethical and regulatory concerns exert a statistically significant negative effect on perceived educational quality. Respondents who report higher levels of concern regarding academic integrity, data protection, and insufficient institutional regulation tend to assess educational quality less favorably. This finding highlights a critical tension in artificial intelligence integration: while technological usage enhances perceived quality, unresolved ethical issues may partially offset these gains. The result reinforces arguments in the literature that trust, transparency, and accountability are central determinants of successful artificial intelligence adoption in education.

The explanatory power of the regression model, which accounts for approximately 28 percent of the variation in perceived higher education quality, is substantial for survey-based educational research. This indicates that artificial intelligence usage and ethical perceptions together represent key drivers of quality assessments, even when individual-level heterogeneity is present. However, the remaining unexplained variation suggests that artificial intelligence should be understood as one component of a broader quality ecosystem, rather than as a standalone solution.

From an institutional perspective, the UNEC case illustrates that artificial intelligence adoption is already occurring at a moderate level, primarily through individual and informal usage rather than fully institutionalized strategies. The strong positive effect of usage frequency implies that further quality improvements may be achievable through more systematic integration of artificial intelligence tools into curricula and teaching practices. At the same time, the negative role of ethical concerns indicates that such integration must be accompanied by clear institutional rules, guidance from instructors, and transparent governance frameworks.

Overall, the discussion of results suggests that the impact of artificial intelligence on higher education quality is conditional rather than automatic. The benefits of artificial intelligence materialize when usage is widespread and meaningful, but they are constrained when ethical uncertainty and regulatory ambiguity persist.

## 4 Conclusion

This study examined the relationship between artificial intelligence usage and higher education quality using original survey data from the Azerbaijan State University of Economics. Based on real empirical evidence, the findings demonstrate that artificial intelligence usage is positively and significantly associated with perceived higher education quality, while ethical and regulatory concerns exert a negative influence.

The results indicate that artificial intelligence enhances educational quality primarily by supporting instructional effectiveness, facilitating self-directed learning, and enabling deeper integration of digital tools into the learning process. However, these benefits are not unconditional. Ethical concerns related to academic integrity, data privacy, and insufficient regulation reduce perceived quality gains and may undermine trust in artificial intelligence-driven educational practices.

From a policy and managerial perspective, the findings carry several important implications. First, higher education institutions should recognize artificial intelligence as a strategic quality-enhancing resource rather than a marginal technological innovation. Second, expanding artificial intelligence usage alone is insufficient; universities must simultaneously invest in ethical governance, regulatory clarity, and instructor guidance. Third, institutional efforts to strengthen digital competence among students and academic staff are likely to amplify the positive effects of artificial intelligence adoption.

In the context of developing and transition economies, the UNEC case demonstrates that artificial intelligence can contribute meaningfully to higher education quality even in resource-constrained environments. However, sustainable quality improvements require coordinated institutional action that aligns technological adoption with ethical standards and quality assurance mechanisms.

By providing real, survey-based empirical evidence from a non-Western higher education system, this study contributes to a more balanced and globally relevant understanding of the role of artificial intelligence in higher education quality. Future research should build on these findings by employing longitudinal designs, multi-institutional samples, and mixed-method approaches to further explore the long-term implications of artificial intelligence integration in higher education systems.

Despite its contributions, this study is subject to several limitations. First, the analysis relies on self-reported survey data, which may be influenced by subjective perceptions and response bias. Future studies could complement survey-based measures with objective performance indicators, such as learning outcomes, completion rates, or administrative efficiency metrics.

Second, the research adopts a cross-sectional design, limiting the ability to draw causal inferences about the long-term impact of AI adoption on educational quality. Longitudinal studies would provide valuable insights into how the effects of AI evolve over time as institutional experience and governance frameworks mature.

Third, the empirical scope is limited to a single institution, which may constrain the generalizability of the findings. While the case study approach enables in-depth contextual analysis, future research could employ multi-institutional or cross-country designs to compare AI adoption patterns and quality outcomes across different higher education systems.

Finally, future research could explore discipline-specific effects of AI adoption, as well as 0the differential impacts on academic staff and students. Qualitative approaches, including interviews and focus groups, may also enrich understanding of how ethical concerns, trust, and institutional culture shape the effectiveness of AI integration in higher education.

## References

1. Siemens, G., & Baker, R. Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK 2012) (pp. 252–254). (2012). https://doi.org/10.1145/2330601.2330661
2. UNESCO. Artificial intelligence and education: Guidance for policymakers. UNESCO Publishing. (2022). https://unesdoc.unesco.org/

3. Binns, R. Fairness in machine learning. Communications of the ACM, 61(11), 40–45. (2018). https://doi.org/10.1145/3278721
4. World Economic Forum. Shaping the future of learning: The role of AI in Education 4.0. (2024). https://www.weforum.org/
5. VanLehn, K. The relative effectiveness of intelligent tutoring systems: A meta-analysis. Educational Psychologist, 46(4), 197–221. (2011). https://doi.org/10.1080/00461520.2011.611369
6. Schneider, M., Dorn, E., Huber, C., & Sellen, P. Personalized learning with artificial intelligence. Computers & Education, 142, Article 103636. (2019). https://doi.org/10.1016/j.compedu.2019.103636
7. Baillifard, A., Gabella, M., Lavenex, P. B., & Martarelli, C. S. Implementing learning principles with a personal AI tutor: A case study. arXiv. (2023). https://doi.org/10.48550/arXiv.2309.13060
8. Alam, G. F., Wiyono, B. B., Burhanuddin, Muslihati, & Mujaidah, A. Artificial intelligence in education world: Opportunities, challenges, and future research recommendations. Fahima: Jurnal Pendidikan dan Kajian Keislaman, 3(2), 1–15. (2024).
9. Ifenthaler, D. Learning analytics. TechTrends, 59(3), 15–21. (2015). https://doi.org/10.1007/s11528-015-0852-5
10. Alfonseca, E., & Pérez-Marín, D. Automatic assessment of open-ended questions. Educational Technology & Society, 14(1), 100–112. (2011).
11. Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. Computers & Education, 138, 62–75. (2019). https://doi.org/10.1016/j.compedu.2019.04.008
12. Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. Current status, opportunities and challenges of augmented reality in education. Computers & Education, 62, 41–49. (2020). https://doi.org/10.1016/j.compedu.2012.10.024
13. World Economic Forum. Shaping the future of learning: The role of AI in Education 4.0. (2024). https://www.weforum.org/
14. Tikhomirova, O. Artificial intelligence in higher education: Institutional challenges and perspectives. Higher Education in Russia, 30(4), 45–58. (2021).
15. Kameneva, N. A. Использование искусственного интеллекта в высшем образовании [The use of artificial intelligence in higher education]. In DHTE 2024 Conference Proceedings (pp. 374–386). (2024).

# ICT in Pattern Recognition section

# Ensemble Strategies for Turkish Legal Extractive Summarization Using Semantic Voting and ROUGE-Weighted Averaging

Maha ALBAYATI[1][0000−0002−4259−9168] and Oğuz FINDIK[2][0000−0001−5069−6470]

[1] Karabuk University, Karabuk 78000, Turkey
[2] Karabuk University, Karabuk 78000, Turkey
mahaa6022@gmail.com
oguzfindik@karabuk.edu.tr
https://www.karabuk.edu.tr/en

**Abstract.** The growing volume of Turkish legal documents poses an important challenge for legal professionals who must efficiently review complex, domain-specific texts. Although extractive summarization has shown promise across many languages, Turkish remains difficult due to its agglutinative morphology and the rigid structure of legal discourse. Existing approaches mainly rely on single methods and often fail to capture both semantic relevance and contextual structure. This study examines whether ensemble strategies can improve the extractive summarization of Turkish legal documents beyond traditional baseline methods. TF-IDF, TextRank, and a BERTurk-Legal-based extractive ranker serve as baseline methods, whereas Semantic Voting and ROUGE-guided weighting are evaluated as ensemble extensions built on these baselines. A curated dataset of 100 Turkish court decisions was compiled and annotated with manually prepared extractive reference summaries. Performance was assessed with ROUGE metrics. Semantic Voting obtained ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4681, 0.3632, and 0.3968, respectively. ROUGE-Weighted Averaging obtained the highest scores, reaching 0.6423, 0.6377, and 0.6423 on ROUGE-1, ROUGE-2, and ROUGE-L. Overall, the findings suggest that the ensemble configuration, particularly ROUGE-Weighted Averaging, improves summary quality relative to the traditional baselines in this experimental setting.

**Keywords:** Ensemble methods · Extractive summarization · ROUGE evaluation · Semantic similarity · Turkish legal summarization

## 1 Introduction

Natural Language Processing (NLP) is a branch of artificial intelligence that enables machines to process and analyze human language. It is used in many tasks, including machine translation, sentiment analysis, information retrieval, question answering, and summarization. Text summarization condenses a document while

preserving its essential information [1]. Approaches are generally divided into extractive summarization, which selects original sentences [3, 13], and abstractive summarization, which paraphrases ideas in new wording [2, 7]. Because extractive methods preserve the original legal wording, they are particularly suitable for domains that require fidelity to the source language, such as Turkish court decisions. Classic extractive approaches rely mainly on statistical and graph-based evidence. The canonical traditional baselines are TF-IDF, which ranks sentences by the importance of their constituent terms, and TextRank, which propagates importance through a sentence-similarity graph [16]. To strengthen the comparison setting, this study also includes a modern neural extractive baseline based on the legal-domain transformer checkpoint BERTurk-Legal [17]. Against this background, the paper examines ensemble-based extensions of traditional methods. TF-IDF, TextRank, and BERTurk-Legal are treated as the baseline systems, while Semantic Voting and ROUGE-Weighted Averaging are evaluated as two ensemble strategies that combine or re-rank their outputs. On a corpus of 100 judgments, Semantic Voting reaches ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4681, 0.3632, and 0.3968, whereas ROUGE-Weighted Averaging attains 0.6423, 0.6377, and 0.6423. These findings suggest that ensemble-based score fusion can improve on the underlying baselines in this experimental setting. The contribution of this study is primarily domain-oriented and empirical, focusing on controlled comparisons and evaluations of Turkish legal extractive summarization rather than on proposing a fundamentally new summarization architecture. Future work may extend the dataset beyond the present 100-case corpus and examine broader cross-validation settings or additional fusion strategies.

## 2 Problem Statement

Turkish summarization research has evolved along two broad tracks: general-domain work, mainly based on news corpora, and a smaller body of studies in specialized domains such as academic writing, biomedicine, social media, and, more recently, law. This development helps explain why Turkish legal summarization remains underexplored [11, 12]. Early work on Turkish summarization relied largely on surface-level statistics. TF-IDF weighting, Latent Semantic Analysis, and PageRank-style methods were applied to Turkish texts and news-style documents [4, 3]. As larger Turkish summarization resources became available, research gradually moved from lexical heuristics toward neural and sequence-to-sequence approaches [5, 7]. VBART, a Turkish sequence-to-sequence language model trained from scratch on a large Turkish corpus, reported competitive results for Turkish generation and summarization tasks [6]. Research outside mainstream news has also explored several specialized settings. In academic writing, researchers have examined deep learning approaches for summarizing Turkish scientific papers [8]. Biomedical work has investigated extractive summarization for Turkish medical papers [9]. Social-media studies remain exploratory, with recent work examining summary informativeness in multilingual

tweet collections under noisy conditions [10]. However, these studies focus on domains with less rigid structure and less formal terminology than legal texts. Legal-domain studies remain limited. A hybrid NLP-based study on Turkish Constitutional Court decisions reported only modest gains over a TF-IDF baseline [11]. A later transformer-based pipeline generated multi-document civil-case summaries, but it relied on a single model and did not examine score-level ensemble fusion [12]. Published work, therefore, provides only limited evidence on how complementary extractive signals and sentence-level semantics can be combined and evaluated for Turkish legal summarization. Early ensemble work combined centroid-based extractors in English news domains [13], and later studies explored score aggregation across multiple summarizers [14]. More recent systems based on pretrained contextual encoders further improved the quality of extractive ranking [15]. Against this background, we examine whether ensemble strategies based on score fusion and semantic re-ranking can provide measurable improvements over TF-IDF and TextRank for Turkish court decisions.

## 3 Methodology

This section describes the methodology for improving extractive summaries of Turkish legal documents using ensemble strategies. The experimental framework consists of three baseline models, TF-IDF-based, TextRank-based, and BERTurk-Legal-based extractive summarization, together with two ensemble strategies, Semantic Voting and ROUGE-Weighted Averaging.

### 3.1 TF-IDF-Based Extractive Summarization

In the TF-IDF approach, the document is first preprocessed using the Zemberek NLP toolkit for Turkish. Preprocessing includes tokenization, lowercasing, stopword removal, and cleaning numerical and date-related information. The same preprocessed documents are used consistently throughout the experiments. For each sentence, we compute a score by summing the TF-IDF values of its words, where the TF-IDF score of a word $w$ is calculated as:

$$\text{TF-IDF}(w) = \text{TF}(w) \times \log \frac{N}{\text{DF}(w)} \tag{1}$$

In this equation, $\text{TF}(w)$ is the term frequency of word $w$, representing how often it appears in the current document. $\text{DF}(w)$ is the document frequency, indicating how many documents in the corpus contain the word $w$, and $N$ is the total number of documents. After computing the TF-IDF scores for all words, each sentence is assigned a cumulative score based on the sum of the TF-IDF values of its constituent tokens. The top-3 scoring sentences are then selected as the extractive summary. These summaries are evaluated under the shared ROUGE-based metrics and the common 5-fold cross-validation protocol described in the Experimental Setup.

### 3.2 TextRank-Based Extractive Summarization

TextRank generates a graph-based representation of the document using the same preprocessed sentence sequence, where nodes correspond to sentences and edges are weighted using cosine similarity between sentence TF-IDF vectors. The importance of each sentence $S_i$ is calculated iteratively using the formula:

$$\text{Score}(S_i) = (1 - d) + d \sum_{S_j \in \text{In}(S_i)} \frac{\text{Score}(S_j)}{|\text{Out}(S_j)|} \tag{2}$$

In this formula, $d$ is the damping factor, typically set to 0.85. The term $\text{In}(S_i)$ refers to the set of sentences that have an edge pointing to $S_i$ (i.e., incoming links), and $\text{Out}(S_j)$ refers to the sentences that $S_j$ links to (i.e., outgoing links). As illustrated in Figure 1, sentence $S_1$ links to both $S_2$ and $S_3$, so $S_1 \in \text{In}(S_2)$ and $S_1 \in \text{In}(S_3)$. In this case, $\text{Out}(S_1) = \{S_2, S_3\}$. The figure clarifies how incoming and outgoing links influence the flow of importance scores across the graph.

After convergence, the top-3 ranked sentences are selected and reordered to match their original positions in the text for improved coherence. These summaries are evaluated using the same ROUGE-based metrics and 5-fold cross-validation protocol as for the other methods.

In the sentence-similarity graph used by TextRank, the edge weight $\text{sim}(S_i, S_j)$ is symmetric, so every undirected edge appears in both directions. Thus, $\text{In}(S_i)$, the sentences that point to $S_i$, and $\text{Out}(S_i)$, those that $S_i$ points to, contain the same neighbors. We keep the PageRank notation because the sum is taken over incoming neighbors, whereas the denominator $|\text{Out}(S_j)|$ normalizes the vote of each contributing node $S_j$ by its identical out-degree.

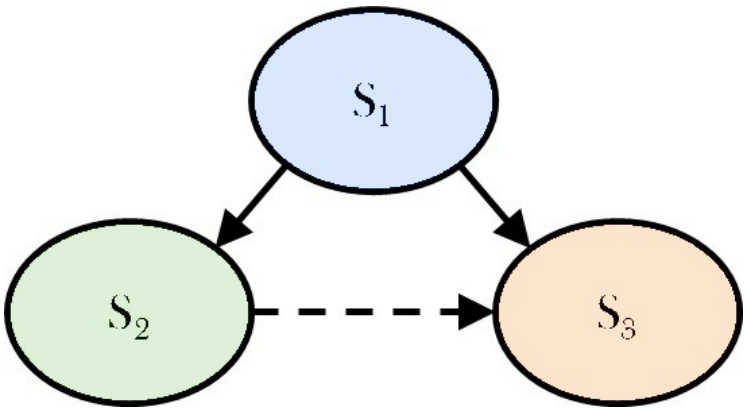



**Fig. 1.** TextRank sentence scoring formula with inbound and outbound links.

### 3.3 BERTurk-Legal-Based Extractive Summarization

To provide a stronger modern neural baseline for Turkish legal summarization, we also evaluated an extractive ranking method built on the domain-adapted transformer checkpoint `KocLab-Bilkent/BERTurk-Legal` [17]. The same preprocessed documents and sentence boundaries were used as in the other methods. For each document, sentence embeddings and a document embedding were obtained from the transformer encoder through mean pooling over the final hidden states. Sentence relevance was then estimated by cosine similarity between each sentence embedding and the corresponding document embedding. The top-3 highest-scoring sentences were selected and reordered according to their original document positions to preserve coherence. BERTurk-Legal was evaluated only as an independent baseline and was not used as an input source in the Semantic Voting or ROUGE-Weighted Averaging ensembles. This baseline keeps the extractive setting explicit while testing whether a legal-domain-pretrained encoder offers an advantage over traditional lexical and graph-based baselines. The resulting summaries were evaluated using the same ROUGE metrics and 5-fold cross-validation protocol used throughout the study.

### 3.4 Semantic Voting Ensemble

The semantic voting strategy merges and re-ranks sentences based on their semantic closeness to the original document. It operates on the same preprocessed documents used by the baseline methods. First, candidate sentences are collected from TF-IDF and TextRank summaries, and duplicates are removed. The full document and each candidate sentence are encoded using the SentenceTransformer model `paraphrase-multilingual-MiniLM-L12-v2`. Then, cosine similarity is computed between each document embedding and the corresponding sentence embedding. Sentences are ranked based on these similarity scores, and the top-3 most semantically relevant ones are selected and ordered according to their original position in the document. This ensemble construction step keeps sentence selection explicit while producing summaries that are evaluated under the same ROUGE metrics and 5-fold cross-validation protocol.

### 3.5 ROUGE-Weighted Averaging Ensemble

In the weighted averaging strategy, the contributions of TF-IDF and TextRank models are dynamically balanced based on validation-set ROUGE scores rather than test-set results. The method uses the same preprocessed documents and preserves explicit sentence selection through score fusion. To avoid evaluation leakage, the data used for weight tuning are explicitly separated from the data used for final reporting. Within each outer cross-validation run, TF-IDF and TextRank are first executed on an internal validation subset, and their average ROUGE-1, ROUGE-2, and ROUGE-L F1-scores on that subset are computed. These validation scores are then normalized to obtain fixed weights $w_1$ and $w_2$ such that:

$$w_1 + w_2 = 1 \tag{3}$$

For each document, candidate sentences are defined as the union of the top-ranked sentences returned by the TF-IDF and TextRank base summarizers. Each candidate sentence then keeps its original sentence-level TF-IDF score and its original sentence-level TextRank score. The TextRank component is therefore taken directly from the graph-based ranking scores described in Eq. (2), not from sentence length or any other proxy. Because the two scoring functions operate on different numeric scales, both score vectors are min-max normalized at the document level before fusion. After the weights are fixed on the internal validation subset, the same weights are transferred unchanged to the held-out test subset of that fold. A final score for each candidate sentence is calculated as:

$$\text{FinalScore}(S_i) = w_1 \times \widehat{\text{TF-IDF}}(S_i) + w_2 \times \widehat{\text{TextRank}}(S_i) \tag{4}$$

The candidate sentences are then ranked by their final scores, and the top-3 highest-scoring sentences are selected as the final extractive summary. The selected sentences are reordered according to their original document positions to preserve coherence. This formulation keeps the weighted ensemble faithful to the actual TF-IDF and TextRank sentence rankings while preserving a leakage-free protocol, since the ROUGE-derived weights are tuned only on the internal validation subset of each fold and never updated with information from the held-out test documents. The resulting summaries are evaluated using the same ROUGE metrics and 5-fold cross-validation protocol used throughout the study.

## 4 Experimental Setup

### 4.1 Dataset

We constructed a custom dataset of 100 Turkish legal documents, each representing an actual court decision or judgment. The documents were manually cleaned to remove metadata, formatting inconsistencies, dates, and case numbers. The cleaned text consists primarily of the main legal reasoning and verdict sections, making it suitable for extractive summarization. Each document was paired with a manually prepared extractive reference summary composed of three source sentences. The annotation procedure followed a fixed guideline: one sentence to represent the court decision, one to capture the main legal issue or reasoning, and one to state the final outcome. After the initial sentence selection, each reference summary was checked for fidelity to the source wording, redundancy, and coverage of the core legal content before being included in the benchmark. Formal inter-annotator agreement was not measured in the current study. However, the reference summaries were prepared under a fixed annotation guideline and manually verified for fidelity, redundancy, and role coverage. Future work should incorporate multi-annotator validation to further strengthen annotation reliability.

The corpus size was kept moderate for practical reasons. Each document required manual cleaning, document-level review, and manual preparation of an extractive reference summary, and these steps were applied to legally sensitive

texts. Accordingly, the dataset should be regarded as a curated benchmark for controlled methodological comparison rather than as a large-scale corpus intended for broad statistical coverage.

### 4.2 Experimental Protocol

To ensure a leakage-free evaluation, we adopted document-level 5-fold cross-validation. In each outer run, one fold (20% of the corpus) was reserved as a held-out test fold, while the remaining four folds (80%) formed the development portion for that run. The development portion was then split internally into a development subset used for pipeline preparation and an internal validation subset. Because TF-IDF, TextRank, and the BERTurk-Legal baseline are unsupervised extractive methods, the development subset was used only to execute the base pipelines under the current fold configuration, whereas the validation subset was used exclusively to tune the ROUGE-based weights of the weighted ensemble. Semantic Voting was applied within the same outer folds without additional parameter tuning beyond the fixed embedding model. Once these weights were fixed for a given run, no further tuning was performed. The learned weights were then transferred unchanged to the held-out test fold, and the process was repeated until every fold had served as the outer test fold once. All five methods generated three-sentence summaries from the same preprocessed documents and were evaluated with the same ROUGE metrics on the held-out folds. Final ROUGE values reported in this paper correspond to the mean scores across the five held-out test folds. This nested protocol clearly separates development, validation, and final evaluation and prevents evaluation leakage.

### 4.3 Preprocessing

The preprocessing pipeline was implemented using the Zemberek NLP library, which supports Turkish morphology and sentence extraction. For each document, tokenization was applied using Zemberek's `TurkishTokenizer`, and sentence boundaries were detected using the `TurkishSentenceExtractor`. The text was further normalized by converting all characters to lowercase and removing numeric expressions, legal article numbers, and common functional patterns that act as legal-domain stopwords. The resulting standardized text served as input to the TF-IDF and TextRank summarization pipelines.

### 4.4 Summarization Pipeline

Each sentence's importance in the TF-IDF summarization method was calculated as the sum of the TF-IDF scores of its tokens, considering only terms present in the current document. The top-scoring sentences were then selected to form the extractive summary. The TextRank summarization method constructed a similarity graph in which each node represented a sentence, and edges were weighted by cosine similarity between the TF-IDF vectors. Sentence importance

was computed using the TextRank algorithm with a damping factor of 0.85, and the highest-ranking sentences were selected to build the summary. The BERTurk-Legal baseline encoded both full documents and their candidate sentences with a legal-domain transformer model, ranked sentences by document–sentence cosine similarity, and selected the top-3 sentences in their original document order.

Two ensemble strategies were applied to refine the TF-IDF and TextRank-based summaries. The semantic voting ensemble re-ranked candidate sentences, collected from both TF-IDF and TextRank outputs, based on their semantic similarity to the entire document using multilingual sentence embeddings. The ROUGE-weighted averaging ensemble computed dynamic weights using only the internal validation subset of each outer fold and then applied these fixed weights during testing so that held-out test documents did not contribute to weight estimation. In other words, BERTurk-Legal remained a standalone baseline for comparison, whereas the two ensemble models operated only on TF-IDF and TextRank candidate sentences and sentence-level scores. Across all five models, preprocessing, sentence selection, ensemble construction where applicable, evaluation metrics, and the 5-fold cross-validation protocol were kept explicit and consistent. Figure 2 summarizes the overall experimental pipeline, while Figure 3 shows the model-specific summarization workflow for TF-IDF, TextRank, BERTurk-Legal, Semantic Voting, and ROUGE-Weighted Averaging.

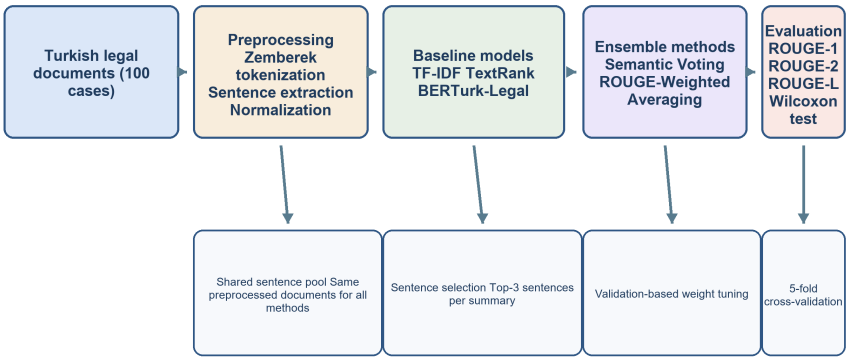



**Fig. 2.** Overall experimental pipeline. The figure shows the shared preprocessing stage, baseline generation with TF-IDF, TextRank, and BERTurk-Legal, ensemble construction, evaluation metrics, and the common 5-fold cross-validation protocol. BERTurk-Legal is evaluated only as a standalone baseline, whereas the ensemble models operate on TF-IDF and TextRank outputs.

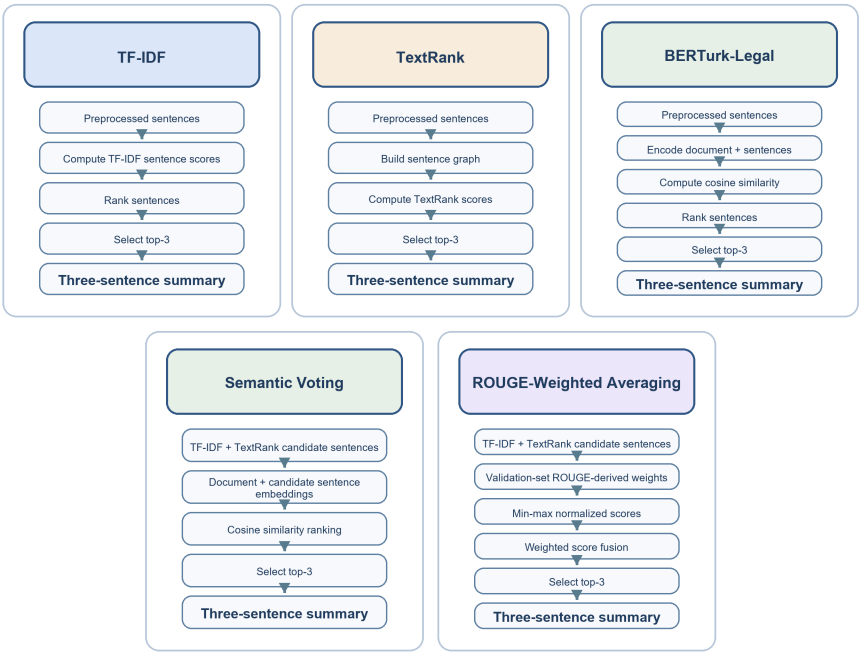



**Fig. 3.** Model-specific summarization workflows. TF-IDF, TextRank, and BERTurk-Legal produce baseline summaries directly, whereas Semantic Voting re-ranks candidate sentences through document–sentence cosine similarity, and ROUGE-Weighted Averaging combines validation-derived weights with min-max normalized baseline scores.

### 4.5 Evaluation Metrics

All generated summaries were evaluated against reference summaries using the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric [18]. The evaluation focused on three key variants: ROUGE-1, which measures the overlap of unigrams (individual words); ROUGE-2, which captures the overlap of bigrams (two-word sequences); and ROUGE-L, which identifies the longest common subsequence between generated and reference summaries. For each metric, F1 scores were reported. In addition to the individual scores, we also computed a ROUGE sum metric, defined as the cumulative sum of ROUGE-1, ROUGE-2, and ROUGE-L F1 scores, to provide a consolidated measure of overall summarization performance. The reported averages were computed on the five held-out test folds only, whereas the ROUGE-weighted ensemble parameters were estimated separately from the internal validation subset inside each fold. To complement the mean scores, we also applied paired Wilcoxon signed-rank testing to the document-level ROUGE outputs collected from the held-out test folds to assess whether the observed differences between the ensemble models and the strongest baseline were statistically reliable. Statistical significance was assessed at the $p < 0.05$ level.

### 4.6 Reproducibility Statement

The experimental pipeline is described in sufficient detail to support methodological transparency, including preprocessing, sentence selection, ensemble construction, evaluation metrics, and the 5-fold cross-validation protocol. The implementation details reported in the manuscript are sufficient to support methodological replication of the experimental design and its evaluation procedure. The original dataset is not publicly released due to legal and institutional constraints governing the source documents.

**Table 1.** Mean ROUGE scores for summarization models across 5-fold cross-validation.

| **Approach name** | **Model name** | **R_1** | **R_2** | **R_L** | **R_SUM** |
|---|---|---|---|---|---|
| Baseline models | TF-IDF | 0.4879 | 0.3940 | 0.4303 | 1.3122 |
| | TextRank | 0.4761 | 0.3817 | 0.4227 | 1.2805 |
| | BERTurk-Legal | 0.4410 | 0.3217 | 0.3796 | 1.1424 |
| Ensemble models | Semantic Voting | 0.4681 | 0.3632 | 0.3968 | 1.2281 |
| | ROUGE-Weighted Averaging | **0.6423** | **0.6377** | **0.6423** | **1.9223** |

## 5 Results and Discussion

We evaluated three base models (TF-IDF, TextRank, and BERTurk-Legal) together with two ensemble approaches (Semantic Voting and ROUGE-Weighted

Averaging) using ROUGE-1 (R1), ROUGE-2 (R2), ROUGE-L (RL), and the cumulative score (RSUM). The detailed results are reported in Table 1. The table presents mean scores over the five held-out test folds, while the ROUGE-based weights used by the weighted ensemble were tuned only on the internal validation subset of each outer fold. In addition to the mean scores, paired Wilcoxon signed-rank tests were applied to the held-out ROUGE values at the document level to reduce the risk of over-interpreting partition-specific variation. Among the single-model baselines, TF-IDF slightly outperformed TextRank, achieving ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4879, 0.3940, and 0.4303, with a cumulative RSUM of 1.3122. TextRank obtained 0.4761, 0.3817, and 0.4227, with an RSUM of 1.2805. The BERTurk-Legal baseline reached 0.4410, 0.3217, and 0.3796, with a cumulative RSUM of 1.1424. The difference among the three baselines, therefore, remains clear, and TF-IDF remains the highest-scoring single-model reference under the present evaluation setup.

Figure 4 visualizes the same ROUGE comparison and clearly separates the baseline group from the weighted ensemble across all three reported metrics.

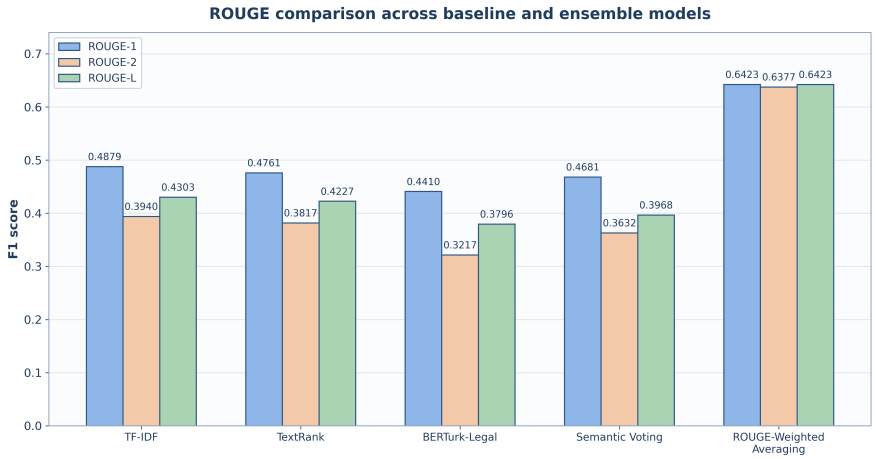



**Fig. 4.** Comparison of ROUGE-1, ROUGE-2, and ROUGE-L F1 scores across the baseline and ensemble models. The grouped bars visually reinforce the stronger overall performance of ROUGE-Weighted Averaging.

The Semantic Voting ensemble reached ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4681, 0.3632, and 0.3968, respectively, with a cumulative RSUM of 1.2281. Under the present configuration, this semantic re-ranking strategy did not surpass the two base models. This suggests that semantic similarity alone was not sufficient to preserve the lexical and structural evidence most relevant to this legal-domain extractive task.

By contrast, the ROUGE-Weighted Averaging ensemble achieved the highest overall scores, with exact ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.6423,

0.6377, and 0.6423, respectively, and a cumulative RSUM of 1.9223. Relative to the strongest baseline, TF-IDF, this corresponds to gains of 31.6% in ROUGE-1, 61.9% in ROUGE-2, 49.3% in ROUGE-L, and 46.5% in RSUM. These findings suggest that score-level fusion of TF-IDF and TextRank yielded higher scores than direct semantic re-ranking in this setting and that the weighted ensemble combines complementary evidence from both base models.

The BERTurk-Legal baseline remains an informative methodologically because it introduces a stronger, modern neural comparator based on a legal-domain-pretrained transformer rather than a purely lexical or graph-based heuristic. However, under the present unsupervised document–sentence similarity formulation, it did not exceed TF-IDF or TextRank. A likely reason is that the checkpoint was originally designed for legal retrieval rather than sentence-level extractive summarization, while the current baseline uses document–sentence cosine similarity without any task-specific fine-tuning. Under these conditions, the model can capture broad legal-domain-relatedness, yet may remain less effective at identifying the most summary-worthy sentences in a short three-sentence extractive output. This pattern suggests that domain adaptation alone was not sufficient to outperform the stronger traditional baselines without additional task-specific training or role-aware supervision.

Overall, these results indicate that the ensemble methods, particularly ROUGE-based weighting, yielded higher scores than the traditional baselines. They also suggest that score-level fusion was more effective in this setting than semantic re-ranking alone.

Several limitations should be noted. The dataset is moderate in size and therefore constrains broad generalization beyond the document types represented in this study, even though it remains suitable for controlled methodological comparison. In addition, formal inter-annotator agreement was not measured in the current study. Future work should expand the corpus and incorporate multi-annotator validation to strengthen annotation reliability and improve the benchmark's generalizability.

## 6 Conclusion

This study examined improvements in Turkish legal extractive summarization using ensemble approaches built on traditional methods. In addition to TF-IDF and TextRank, the evaluation included a BERTurk-Legal-based extractive baseline as a stronger modern neural comparator. Semantic Voting obtained ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4681, 0.3632, and 0.3968, with a cumulative RSUM of 1.2281. In contrast, ROUGE-Weighted Averaging obtained the highest scores, reaching 0.6423, 0.6377, and 0.6423 on ROUGE-1, ROUGE-2, and ROUGE-L, with a cumulative RSUM of 1.9223. Compared with the strongest baseline, TF-IDF, the weighted ensemble improved ROUGE-1 by 31.6%, ROUGE-2 by 61.9%, ROUGE-L by 49.3%, and RSUM by 46.5%. These findings suggest that the ensemble strategy provided the strongest overall performance among the methods considered in this setting. The proposed methods

are lightweight and may be suitable for integration into practical legal NLP workflows.

Future work may explore several extensions. First, incorporating deep learning models such as BERT- or LSTM-based classifiers into the ensemble may improve semantic understanding and contextual relevance. Second, experiments on larger datasets and additional legal document types, such as contracts, regulations, or administrative decisions, would help assess generalizability more fully. Human evaluation by legal experts would also provide useful qualitative evidence. Finally, combining extractive and abstractive components may support the development of hybrid summarization systems for the legal domain.

## References

1. Yang, B., Luo, X., Sun, K., Luo, M.Y.: Recent progress on text summarisation based on BERT and GPT. In: *Natural Language Processing and Information Systems*, LNAI, vol. 14120, pp. 225–241. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-40292-0_19
2. Alipour, N., Aydın, S.: Abstractive summarization using multilingual text-to-text transfer transformer for the Turkish text. *IAES International Journal of Artificial Intelligence (IJ-AI)* **14**(2), 1587–1596 (2025). https://doi.org/10.11591/IJAI.V14.I2.PP1587-1596
3. Akülker, E., Tatar, Ç.: Extractive text summarization for Turkish: Implementation of TF-IDF and PageRank algorithms. In: *Proceedings of the International Symposium on Intelligent Systems*, LNNS, vol. 544, pp. 688–704. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-16075-2_51
4. Ozsoy, M.G., Cicekli, I., Alpaslan, F.N.: Text summarization of Turkish texts using latent semantic analysis. In: *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pp. 869–876. Coling 2010 Organizing Committee, Beijing, China (2010). `https://aclanthology.org/C10-1098/`
5. Baykara, B., Güngör, T.: Abstractive text summarization and new large-scale datasets for agglutinative languages: Turkish and Hungarian. *Language Resources and Evaluation* **56**(3), 973–1007 (2022). https://doi.org/10.1007/S10579-021-09568-Y
6. Türker, M., Ari, M.E., Han, A.: VBART: The Turkish LLM. *arXiv preprint* arXiv:2403.01308 (2024). `https://arxiv.org/abs/2403.01308`
7. Baykara, B., Güngör, T.: Turkish abstractive text summarization using pretrained sequence-to-sequence models. *Natural Language Engineering* **29**, 1275–1304 (2023). https://doi.org/10.1017/S1351324922000195
8. Alagoz, N.K., Kucuksille, E.U.: Extractive summarization of scientific papers in Turkish with deep learning. *Research Square preprint* (2022). https://doi.org/10.21203/RS.3.RS-2292145/V1
9. Kuş, A., Acı, Ç.I.: An extractive text summarization model for generating extended abstracts of medical papers in Turkish. *Bilgisayar Bilimleri ve Teknolojileri Dergisi* **4**(1), 19–26 (2023). https://doi.org/10.54047/bibted.1260697
10. Mahindrakar, S.M., Mondal, T., Arosh, S., Dhakne, A., Chavan, D.D.: Exploration of summary informativeness and topic richness through mining multilingual tweets: A study on Turkey earthquake 2023. *Social Network Analysis and Mining* **14**(1), 117 (2024). https://doi.org/10.1007/S13278-024-01386-8

11. Turan, T., Kucuksille, E.U.: Summarization, prediction, and analysis of Turkish Constitutional Court decisions with explainable artificial intelligence and a hybrid natural language processing method. *IEEE Access* **13**, 59766–59779 (2025). https://doi.org/10.1109/ACCESS.2025.3556725
12. Albayati, M.A.A., Findik, O.: A hybrid transformer-based framework for multi-document summarization of Turkish legal documents. *IEEE Access* **13**, 37165–37181 (2025). https://doi.org/10.1109/ACCESS.2025.3545750
13. Radev, D.R., Jing, H., Budzikowska, M.: Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies. In: *Proceedings of the NAACL-ANLP Workshop on Automatic Summarization*, pp. 21–30, Seattle, WA, USA (2000)
14. Umejiaku, A.P., Zhou, F., Sheng, V.S.: Ensemble summarization models to leverage performance. In: *5th International Conference on Computing and Big Data (ICCBD)*, pp. 56–61. IEEE (2022). https://doi.org/10.1109/ICCBD56965.2022.10080715
15. Liu, Y., Lapata, M.: Text summarization with pretrained encoders. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3730–3740, Hong Kong, China (2019). https://doi.org/10.18653/v1/D19-1387
16. Mihalcea, R., Tarau, P.: TextRank: Bringing order into text. In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 404–411. Association for Computational Linguistics, Barcelona, Spain (2004). `https://aclanthology.org/W04-3252/`
17. Öztürk, C.E., Özçelik, Ş.B., Koç, A.: A transformer-based prior legal case retrieval method. In: *2023 31st Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4. IEEE, Istanbul, Turkey (2023). https://doi.org/10.1109/SIU59756.2023.10223938
18. Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries. In: *Text Summarization Branches Out*, pp. 74–81. Association for Computational Linguistics, Barcelona, Spain (2004). `https://aclanthology.org/W04-1013/`

# Lexicon-Augmented Explainable Aspect-based Sentiment Analysis for Azerbaijani Finance and Business Text

Yusuf Evren Aykaç[1,2], Merve Özkan[1], and Refik Samet[1]

[1] Ankara University, Ankara, Türkiye
[2] Ankara Yıldırım Beyazıt University, Ankara, Türkiye
yeaykac@aybu.edu.tr

**Abstract.** Aspect-based sentiment analysis (ABSA) for Azerbaijani has received limited attention, and finance / business text introduces domain terms whose polarity can be context-dependent. We study a Finance / Business Azerbaijani ABSA dataset with ∼27.2K aspect-level instances (≈18 tokens on average) and examine whether the SentiAzNet polarity lexicon can complement a multilingual transformer. Using XLM-RoBERTa as the base encoder, we integrate lexicon cues in two lightweight ways: (1) concatenating token-level polarity features and (2) adding an attention bias towards lexicon-matched tokens. We also define SentiAzNet-Fin through a small set of finance-specific polarity overrides. Lexicon coverage is sparse (3.17% of tokens; ∼38% of instances have ≥1 hit), so the lexicon acts as an optional cue rather than a primary signal. Across 5-fold cross-validation, the attention-bias model with SentiAzNet-Fin reaches 0.769 accuracy and 0.738 macro-F1, improving over vanilla XLM-R by 3.6 macro-F1 points (McNemar $p < 0.001$). For analysis, we compare SHAP attributions, attention saliency, and lexicon hits; agreement between these signals tends to coincide with higher correctness. We further summarize frequent errors such as sarcasm, neutral–negative ambiguity, mixed sentiment, and code-switching.

**Keywords:** Aspect-based sentiment analysis · XLM-RoBERTa · financial sentiment analysis · sentiment lexicon · explainable AI · SHAP · attention

## 1 Introduction

Finance and business text is often short and mixed in tone: a single comment may praise *service* but complain about *fees*, or simply report a rate change without an explicit stance. In these settings, sentence-level sentiment is frequently too coarse. Aspect-Based Sentiment Analysis (ABSA) aims to address this by predicting sentiment with respect to a given aspect term or category rather than assigning one label to the whole sentence.

Most ABSA benchmarks and modeling work have focused on high-resource languages, with SemEval shared tasks providing widely used evaluation settings [18, 9, 2]. More recently, multilingual datasets have enabled cross-lingual

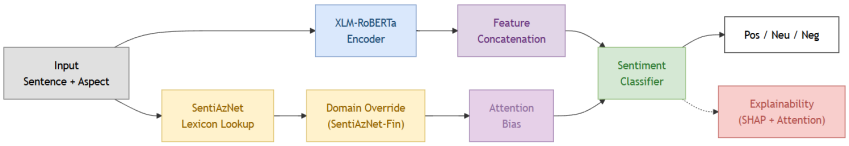



**Fig. 1.** Overview of the proposed ABSA pipeline with lexicon injection and explainability analysis.

evaluation [23]. Still, for low-resource languages, aspect-level supervision and domain-specific resources are often limited, and cross-lingual transfer does not fully remove data and domain gaps [12, 19].

Azerbaijani is a low-resource Turkic language where existing sentiment work has mostly targeted sentence-level classification in news and social media [10]. In finance and business, the vocabulary is more specialized, and polarity can be context-dependent; similar issues are well known in English financial NLP and motivated domain lexicons such as Loughran–McDonald [14] and domain-adapted encoders such as FinBERT [3]. While multilingual transformers like XLM-R are a strong starting point [7], neutral labels, sarcasm, and occasional code-switching still create practical difficulties [21].

In this paper, we study a Finance/Business Azerbaijani ABSA dataset with ∼27.2k aspect-level instances and ask whether an external polarity lexicon can provide helpful cues. We use SentiAzNet [5] and integrate token-level polarity information into XLM-R in two lightweight ways: feature concatenation and attention bias (Fig. 1). Because lexicon matches are sparse (3.17% of tokens; ∼38% of instances have at least one hit; Table 1), we treat the lexicon as an optional signal rather than something the model can rely on for every instance. We also define a small set of finance-specific polarity overrides (SentiAzNet-Fin) to reduce clear domain mismatches.

Finally, we use explainability analysis mainly to understand model behavior and failure cases. We compute SHAP attributions [15] and inspect attention patterns, while keeping in mind that attention weights are not guaranteed to be faithful explanations by themselves [11, 24].

**Contributions.** We (1) report dataset statistics and lexicon coverage for Finance/Business Azerbaijani ABSA (Table 1, Fig. 2); (2) evaluate two practical lexicon injection strategies for XLM-R and a small domain-adapted variant (SentiAzNet-Fin); (3) provide ablations and significance tests; and (4) include interpretability and error analyses that highlight where the approach helps and where it remains brittle.

## 2 Related Work

This section briefly reviews prior work that motivates our study. Recent surveys summarize the broader sentiment-analysis landscape [1] as well as open problems in cross-lingual ABSA [19], which are directly relevant to our setting. We

organize related work around (i) sentiment resources and transfer for Azerbaijani and other low-resource languages, (ii) financial sentiment analysis where domain lexicons are commonly used, and (iii) ABSA benchmarks and explainability methods.

### 2.1 Azerbaijani Sentiment Resources and Low-Resource Transfer

Azerbaijani sentiment research has mainly focused on sentence-level classification, often on news and social media, and publicly available aspect-level resources are limited [10]. In low-resource settings, multilingual pre-trained language models are commonly used because they can transfer knowledge from high-resource languages. Multilingual BERT (mBERT) [8] and XLM-RoBERTa (XLM-R) [7] are widely used starting points, but transfer quality still depends on domain match and label definitions [12, 19]. Lexicons remain relevant as lightweight priors and as tools for analysis. SentiAzNet, the polarity lexicon for Azerbaijani used in this work, was developed by our research group and has been employed in lexicon-based and hybrid sentiment pipelines [5, 4]; in this paper we build on it but treat it as an external, sparse prior rather than a complete sentiment signal.

### 2.2 Financial Sentiment Analysis and Domain Lexicons

Financial sentiment is often treated separately from general sentiment because domain terms can behave differently in financial contexts. In English, domain lexicons such as Loughran–McDonald were created to account for this mismatch [14]. In parallel, domain-adapted encoders (e.g., FinBERT) and fine-tuned transformer models have been explored for financial text classification [3, 17]. A recurring theme is that dictionary cues can still complement contextual models, particularly for rare or high-impact domain terms, but the benefit depends on coverage and on how well polarity definitions match the target task.

### 2.3 ABSA Benchmarks and Explainability

SemEval ABSA shared tasks defined common evaluation settings for aspect terms and sentiment polarity [18, 9, 2]. Multilingual ABSA datasets have since extended evaluation beyond English [23]. Recent work has also summarized cross-lingual ABSA tasks and open challenges (e.g., domain shift and annotation mismatch) [19], and explored data augmentation with large language models to improve cross-lingual ABSA under limited supervision [20]. While LLM-based and instruction-tuned approaches are increasingly competitive for ABSA, they typically rely on large models and substantial compute or API access at inference time. In contrast, our goal is a frugal, deployable alternative that augments a moderately sized multilingual encoder with a sparse lexical prior, which is attractive in low-resource settings where such resources are constrained. For explainability, SHAP is widely used as a post-hoc attribution method [15]. Attention visualizations are also common, but several studies caution that attention

weights do not necessarily reflect causal importance, motivating careful interpretation and comparisons with other signals [11, 24]. Sarcasm remains a common failure mode for sentiment systems because surface polarity cues can be misleading, and this has motivated sarcasm benchmarks and dedicated modeling work [21].

Next, we describe the dataset and evaluation setting used in our experiments.

## 3 Dataset

*Task and instances.* We focus on three-class aspect-based sentiment classification in the Finance/Business domain. Each label is attached to an aspect term within a sentence and can be written as $\langle s, a_t, a_c, y \rangle$, where $s$ is the sentence, $a_t$ is the aspect term, $a_c$ is the aspect category, and $y \in \{\text{POSITIVE}, \text{NEUTRAL}, \text{NEGATIVE}\}$. Because a sentence may contain multiple aspects with different polarities, we treat each $(s, a_t)$ pair as a separate *aspect-level instance*.

*Data origin and annotation.* We use an in-house Finance/Business Azerbaijani ABSA dataset curated for this study. The data is annotated at aspect level with an aspect term, an aspect category, and a three-class sentiment label (POSITIVE, NEUTRAL, NEGATIVE). Annotation was carried out by three native Azerbaijani speakers with a background in linguistics or finance, following written guidelines that defined the aspect categories and provided decision rules for ambiguous and mixed-sentiment cases. Each instance was independently labeled by two annotators, and disagreements were resolved by a third adjudicator. Inter-annotator agreement, measured on a doubly annotated portion comprising 15% of the dataset, was substantial, with a Cohen's $\kappa$ of 0.772 for sentiment polarity and 0.780 for aspect-category assignment. The dataset is currently in-house; we plan to release it, or a representative subset together with the annotation guidelines, to support reproducibility, subject to source and licensing constraints.

*Splits and statistics.* We report descriptive statistics for a representative 70/10/20 train/validation/test partition stratified by sentiment class; model performance is reported with stratified 5-fold cross-validation (Section 5). For lexicon statistics, we report sentence-level match per split and token-level coverage over the full dataset. Table 1 summarizes the split sizes and basic statistics, and Fig. 2 shows the label distribution. Instances are short on average ($\approx$ 18 tokens with the XLM-R tokenizer), but the long tail includes sequences above 120 tokens. Token-level SentiAzNet coverage is computed over the full dataset, while experimental results are obtained using 5-fold cross-validation.

*Lexicon coverage.* To quantify how often external polarity cues are available, we measure overlap with SentiAzNet at token level (*coverage*) and at instance level (*match*: at least one hit). Coverage is sparse in this domain (3.17% of tokens; $\sim$38% of instances contain $\geq$1 hit; Table 1), so lexicon signals are absent for most inputs and are best viewed as optional hints rather than a complete sentiment signal.

**Table 1.** Statistics of the Finance/Business Azerbaijani ABSA dataset for a representative 70/10/20 split.

| Statistic | Train | Validation | Test |
|---|---|---|---|
| Total samples | 19016 | 2716 | 5434 |
| Positive | 8278 | 1183 | 2365 |
| Neutral | 2843 | 406 | 813 |
| Negative | 7895 | 1127 | 2256 |
| Avg. sentence length (tokens) | 18.4 | 18.2 | 18.6 |
| Max sentence length (tokens) | 128 | 125 | 131 |
| Vocabulary size | 42850 | – | – |
| SentiAzNet coverage (token %, overall) | 3.17 | – | – |
| SentiAzNet match (sentence %) | 38.2 | 37.8 | 38.5 |

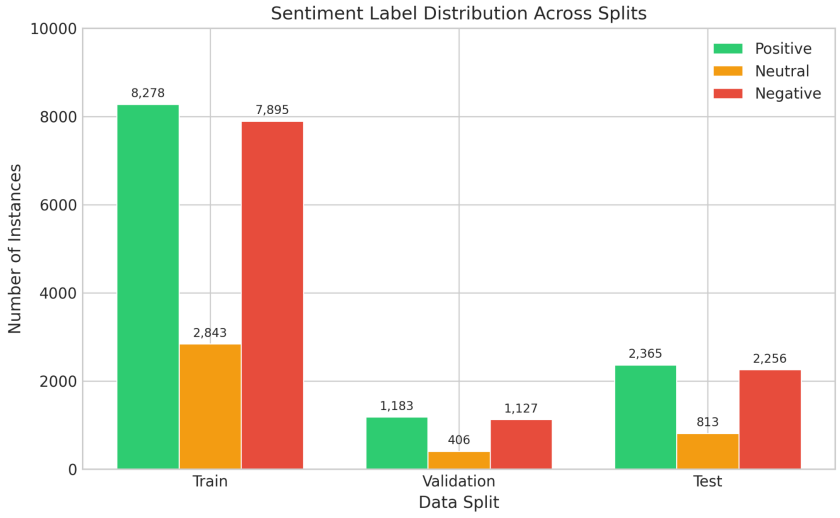



**Fig. 2.** Distribution of sentiment labels in train/validation/test splits.

*Preprocessing.* We apply minimal normalization (whitespace and punctuation) and keep stopwords to preserve negation and discourse markers for explainability. Inputs are tokenized with the XLM-R SentencePiece tokenizer. For lexicon lookup, we match surface words and propagate scores to aligned subwords (details in Section 4).

## 4 Methodology

Figure 1 outlines our approach. For each aspect-level instance, we encode the sentence together with the target aspect using XLM-RoBERTa [7]. In parallel, we compute a token-level polarity signal from the Azerbaijani sentiment lexicon

SentiAzNet [5]. We then inject this signal into the transformer either as an additional input feature (**P1**) or as a soft attention bias (**P2**/**P3**). Finally, we predict one of three sentiment labels (POSITIVE, NEUTRAL, NEGATIVE). We use SHAP and attention patterns mainly for post-hoc analysis (Section 6).

*Aspect-conditioned input.* We follow a standard paired input format where the sentence and the aspect term are concatenated and separated by special tokens: `<s>` tok($s$) `</s>` tok($a_t$) `</s>`, where tok($\cdot$) is the XLM-R SentencePiece tokenization. The encoder produces contextual representations for all tokens. We use the pooled representation of the first token as input to a 3-way classification head.

*Lexicon signal and domain overrides.* SentiAzNet provides prior polarity information for Azerbaijani lexical items. We define a scalar score $l_i$ for each token position $i$. Because XLM-R operates on subword units, lexicon matching is done at the surface-word level and then propagated to the corresponding subword pieces; tokens without a match receive $l_i = 0$. To better match finance/business usage, we define *SentiAzNet-Fin* as a small list of finance-specific polarity overrides. At lookup time, if a word appears in this list, we replace its base polarity with the finance-specific value, resulting in an updated signal $\tilde{l}_i$. We keep this override set small to limit manual effort and to reduce the risk of overfitting to a few terms.

*Lexicon injection via feature concatenation (P1).* In feature concatenation, we treat $\tilde{l}_i$ as an additional input feature and append it to each contextual embedding produced by XLM-R. Since this increases the dimensionality by one, we apply a small linear projection to map the augmented representation back to the original hidden size before passing it to subsequent layers and the classifier. This strategy is easy to implement and serves as a lightweight form of late fusion.

*Lexicon injection via attention bias (P2/P3).* As an alternative, we inject lexicon information inside the self-attention computation, encouraging the model to allocate more attention mass to tokens with stronger polarity cues. Concretely, for scaled dot-product attention [22], we add a lexicon-dependent bias term before the softmax, as shown in Eq. (1):

$$\mathrm{Attn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}} + B_{\mathrm{lex}}\right)V. \tag{1}$$

Here $B_{\mathrm{lex}}$ denotes the lexicon bias term added to the attention scores. We set $B_{\mathrm{lex}}(i, j) = \alpha \cdot m_j$, where $m_j = |\tilde{l}_j|$ and $\alpha$ is a learnable scalar initialized to a small value ($\alpha_0 = 0.1$). Intuitively, this acts as a soft prior: it can help when lexicon cues are present, but the model can still rely on context for the majority of tokens without lexicon matches.

*Training and prediction.* All models are trained with the same objective (cross-entropy over three classes) and the same aspect-conditioned input format. We report results under stratified 5-fold cross-validation and discuss ablation settings in Section 6.

*Explainability signals.* We use SHAP [15] to obtain token-level attributions and extract attention-based saliency from the fine-tuned transformer. Since attention is not guaranteed to be a faithful explanation by itself [11, 24], we treat it as a complementary signal. To summarize agreement, we compare top-$k$ tokens from SHAP and attention with the set of lexicon hits using overlap measures (e.g., Jaccard@$k$) and relate agreement to prediction reliability (Section 6).

## 5 Experimental Setup

*Models.* We compare a lexicon-only baseline (**B1**), classical and neural baselines (**B2**–**B5**), and our lexicon-augmented XLM-R variants (**P1**–**P3**). **B1** is rule-based using only SentiAzNet signals; **B2** is a linear SVM with TF–IDF features; **B3** is a BiLSTM model with FastText embeddings [6]; **B4** fine-tunes mBERT; and **B5** fine-tunes XLM-R without lexicon information. Our proposed models add lexicon cues to XLM-R via feature concatenation (**P1**) or attention bias (**P2**), and **P3** additionally uses the finance-specific polarity overrides (SentiAzNet-Fin).

*Evaluation protocol.* We use stratified 5-fold cross-validation. In each fold, one split ($\approx 20\%$) is held out for testing. From the remaining data we create a small validation split for early stopping and hyperparameter selection, resulting in an overall train/validation/test ratio of approximately 70/10/20 per fold. We report mean scores across folds.

*Metrics.* We report Accuracy, macro-averaged Precision/Recall/F1, and Weighted-F1. Macro-F1 is treated as the primary metric because the class distribution is imbalanced and the neutral class is comparatively difficult.

*Training details.* Transformer-based models are fine-tuned with AdamW [13] and early stopping based on validation macro-F1. Unless stated otherwise, all models share the same preprocessing and tokenization pipeline (Section 3) and the same aspect-conditioned input format (Section 4).

*Significance testing.* For key comparisons we apply McNemar's test [16] on paired test predictions (correct/incorrect per instance), using continuity correction.

*Explainability computation.* We compute token-level SHAP attributions and attention-based saliency scores (Section 4). Since SHAP can be computationally expensive, we compute global summaries on a stratified random subset of 550 test instances (sampled to preserve the test-set class distribution) and report local explanations for selected case studies.

**Table 2.** Main results (5-fold cross-validation).

| Model | Acc. | Macro-P | Macro-R | Macro-F1 | W-F1 |
|---|---|---|---|---|---|
| B1: Rule-based (SentiAzNet only) | 0.412 | 0.358 | 0.341 | 0.349 | 0.398 |
| B2: SVM + TF-IDF | 0.621 | 0.583 | 0.571 | 0.577 | 0.618 |
| B3: BiLSTM + FastText | 0.658 | 0.612 | 0.604 | 0.608 | 0.652 |
| B4: mBERT (fine-tuned) | 0.714 | 0.681 | 0.673 | 0.677 | 0.710 |
| B5: XLM-R (vanilla) | 0.738 | 0.706 | 0.698 | 0.702 | 0.734 |
| P1: XLM-R + SentiAzNet (Feature) | 0.751 | 0.722 | 0.714 | 0.718 | 0.748 |
| P2: XLM-R + SentiAzNet (Attn Bias) | 0.759 | 0.731 | 0.724 | 0.727 | 0.756 |
| **P3: XLM-R + SentiAzNet-Fin (Attn Bias)** | **0.769** | **0.742** | **0.735** | **0.738** | **0.770** |

## 6 Results and Discussion

We begin with a quantitative comparison across baselines and our lexicon-augmented models, and then discuss where the gains come from using class confusions, ablations, and explainability analyses. Finally, we summarize common error patterns that remain challenging in this domain. Some of these remaining difficulties are also discussed in recent surveys of cross-lingual ABSA [19].

### 6.1 Overall Performance

Table 2 summarizes results averaged over 5-fold cross-validation. Among the transformer baselines, XLM-R (**B5**) is slightly stronger than mBERT (**B4**) in our setting. Adding SentiAzNet cues leads to small but consistent gains: feature concatenation (**P1**) improves macro-F1 over **B5**, while attention bias (**P2**) yields a larger improvement. The best configuration (**P3**), which combines attention bias with the finance-specific overrides (SentiAzNet-Fin), reaches 0.769 accuracy and 0.738 macro-F1. Compared to vanilla XLM-R, this is a +3.6 macro-F1 point gain, and the difference is statistically significant under McNemar's test ($p < 0.001$).

### 6.2 Confusions Across Classes

Figure 3(a) shows the row-normalized confusion matrix for **P3**. As expected, the neutral class is the most challenging: 61.3% of neutral instances are predicted correctly, while a substantial portion is predicted as negative (25.0%). Positive and negative classes are recognized more reliably (81.2% and 78.1% on the diagonal), but short factual statements and implicit sentiment still cause confusion, especially near the neutral–negative boundary.

**Table 3.** Ablation study on lexicon injection and training components.

| Configuration | Macro-F1 | ΔF1 (pts) | Accuracy |
|---|---|---|---|
| XLM-R (base) | 0.702 | – | 0.738 |
| + SentiAzNet (Feature concat) | 0.718 | +1.6 | 0.751 |
| + SentiAzNet (Attention bias) | 0.727 | +2.5 | 0.759 |
| + Domain adaptation (SentiAzNet-Fin) | 0.738 | +3.6 | 0.769 |
| – SentiAzNet (bias only) | 0.708 | +0.6 | 0.742 |
| + SentiAzNet + Focal Loss | 0.735 | +3.3 | 0.762 |
| + SentiAzNet + Weighted CE Loss | 0.733 | +3.1 | 0.760 |

### 6.3 Ablation Study

Table 3 breaks down which components contribute most. Feature concatenation provides a modest gain over the base encoder (+1.6 macro-F1 points), while attention bias yields a larger gain (+2.5 points). Adding finance-specific polarity overrides (SentiAzNet-Fin) gives an additional improvement, which is consistent with some polarity cues behaving differently in this domain. The "bias only" ablation (keeping the bias mechanism but removing lexicon information) yields only a small gain, suggesting that the lexicon signal itself matters more than the architectural change alone. We also tried class-imbalance-aware objectives (Focal Loss, weighted cross-entropy); in our setting, they provide improvements close to the best configuration but do not clearly surpass the domain-adapted lexicon variant.

### 6.4 Explainability Signals and Lexicon Alignment

We use explainability mainly to understand model behavior and failure cases, rather than to claim a single "correct" explanation. Figure 3(b) shows a SHAP feature-importance summary for frequently influential tokens in the best model. Many high-importance items are clearly sentiment laden (e.g., *yaxşı* "good", *pis* "bad", *əla* "excellent"), while others reflect common finance/business terms in our corpus (e.g., *komissiya* "fee/commission", *faiz* "interest"). This is broadly consistent with the intuition that opinions in this domain often hinge on service quality, fees, and rate-related language, but SHAP attributions remain context-dependent.

Figure 4 shows a local example. For readability, we visualize token importance over the sentence segment (the aspect segment is omitted in the plot). Attention places most weight on *yaxşıdır* (0.28) and also highlights *komissiya* (0.22) and *çoxdur* (0.20), capturing both the positive and negative cues in the sentence. SHAP attributions align with this picture: *yaxşıdır* contributes positively (+0.34), while *komissiya* and *çoxdur* contribute negatively (-0.28 and -0.22). In our ABSA setting, the final label depends on the target aspect; negative cues are expected to dominate when the aspect relates to fees, while the same sentence could be positive for a service-related aspect.

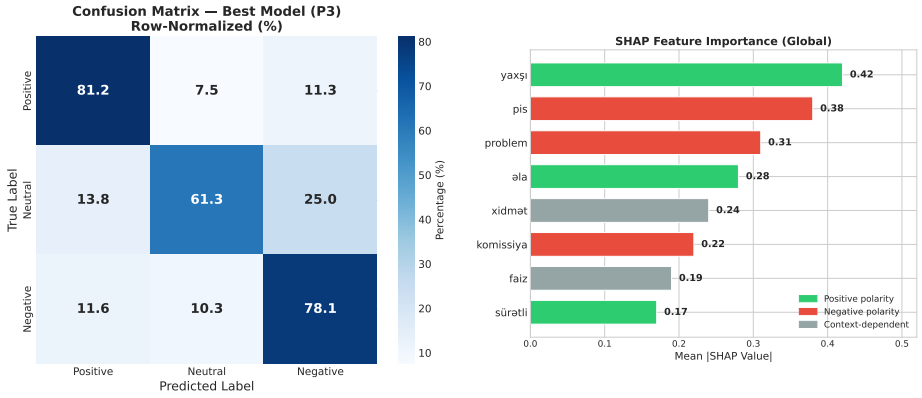



**Fig. 3.** Global analysis for the best model: (a) row-normalized confusion matrix and (b) SHAP feature importance (mean absolute SHAP) for frequently influential tokens.

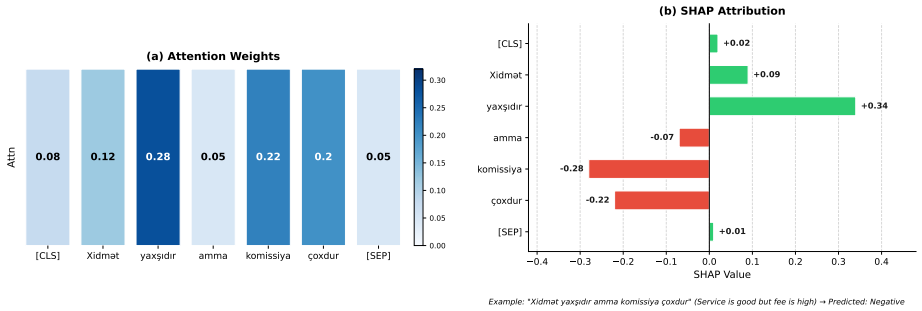



**Fig. 4.** Local explanation example (sentence segment shown): attention weights (left) and SHAP token-level attributions (right) for "Xidmət yaxşıdır amma komissiya çoxdur" (Service is good but fee is high).

To quantify agreement between explanation signals and lexicon hits, Table 4 reports overlap measures. SHAP and attention show moderate overlap (e.g., Jaccard@5 of 0.42), while overlap with lexicon hits is lower. Given sparse lexicon coverage, this is not surprising: the model often needs to rely on contextual cues beyond lexicon-marked tokens. The coverage buckets in the same table show that macro-F1 and average confidence increase with the number of lexicon hits, which is consistent with lexicon cues being most helpful when they are present. At the same time, most test instances (61.8%) have no lexicon hits, so overall performance still depends heavily on contextual representations learned by the encoder. Because these bucket ratios are computed over the pooled test folds in cross-validation, the implied $\geq$1-hit rate (38.2%) is close but not identical to the sentence-level match for the fixed test split in Table 1 (38.5%).

**Table 4.** Alignment between SHAP, attention, and lexicon signals under different lexicon coverage levels.

| Metric | SHAP↔Attn | SHAP↔Lex | Attn↔Lex | 3-way overlap |
|---|---|---|---|---|
| Jaccard@5 | 0.42 | 0.31 | 0.28 | 0.19 |
| Jaccard@10 | 0.51 | 0.38 | 0.34 | 0.24 |
| Kendall's $\tau$ | 0.47 | 0.33 | 0.29 | – |
| **Coverage impact (CV test-fold buckets)** | | | | |
| ≥3 lexicon hits | Macro-F1: 0.812 | Conf.: 0.910 | 8.3% (CV tests) | |
| 1–2 lexicon hits | Macro-F1: 0.756 | Conf.: 0.840 | 29.9% (CV tests) | |
| 0 lexicon hits | Macro-F1: 0.698 | Conf.: 0.730 | 61.8% (CV tests) | |

**Table 5.** Common misclassification patterns (illustrative examples). Frequencies are computed over the 250 manually inspected misclassified instances that could be assigned to one of these categories.

| Error type | Freq. | Example (Az) | True | Pred. |
|---|---|---|---|---|
| Sarcasm / irony | 28.3% | əla bank, hər gün yeni problem tapır | Neg | Pos |
| Neutral↔Negative | 24.1% | Bank faiz dərəcəsini dəyişdi | Neu | Neg |
| Mixed sentiment | 18.7% | Xidmət yaxşıdır amma komissiya çoxdur | Neg | Pos |
| Implicit sentiment | 14.2% | Bu bankdan başqa bank yoxdur ki? | Neg | Neu |
| Code-switching | 8.9% | Çox teşekkür ederim, harika bir hizmet | Pos | Neu |
| Domain jargon | 5.8% | Spread artıb, yield əyrisi düzəlib | Neu | Neg |

### 6.5 Error Analysis

We manually inspected a stratified sample of 250 misclassified instances and grouped them into common patterns (Table 5). Sarcasm/irony is a frequent source of errors, because positive words can be used in clearly negative contexts, which can mislead both lexicon cues and surface-level associations. Neutral–negative ambiguity is also common in finance text: short factual statements (e.g., policy or rate changes) may be annotated as neutral but still carry negative implications for readers. Mixed sentiment and implicit sentiment highlight that cues can be distributed across clauses and may not be explicitly linked to the target aspect. Finally, code-switching and domain-specific jargon remain challenging due to tokenization effects and limited lexicon coverage for borrowed financial terminology.

## 7 Conclusion

We investigated ABSA for Azerbaijani in the Finance/Business domain and examined whether an external polarity lexicon can serve as useful auxiliary information for multilingual transformers. Using a dataset of approximately 27.2k aspect-level instances, we fine-tuned XLM-RoBERTa and injected SentiAzNet

cues via feature concatenation or attention bias. Although lexicon coverage is sparse, lexicon-aware variants generally improved over the vanilla XLM-R baseline, and the strongest results came from combining attention bias with a small set of finance-specific polarity overrides (SentiAzNet-Fin). We also used SHAP attributions, attention saliency, and lexicon hits to analyze model behavior, and found that agreement between these signals often coincides with higher correctness.

**Limitations and future work.** Lexicon coverage is low for finance/business text, so improvements are concentrated in instances where sentiment cues are lexicalized and matched by the lexicon. The dataset used in this study is currently in-house, which limits direct external reproduction; to mitigate this we report detailed split statistics and annotation procedures (Section 3) and intend to release the data, or a representative subset with the annotation guidelines, subject to source and licensing constraints. Persistent error patterns include sarcasm/irony, neutral–negative ambiguity, mixed sentiment, and code-switching. Future work may explore expanding domain-specific lexicon coverage (including loanwords and code-switched terms), incorporating morphology-aware modeling, and adding targeted strategies for sarcasm and pragmatic cues. Another direction is to use weak supervision or data augmentation for cross-lingual ABSA, including recent LLM-based augmentation methods [20]. Human-centered evaluation of explanation usefulness would also help assess the practical value of the proposed analysis.

**Acknowledgments.** This study was funded by TUBITAK (The Scientific and Technological Research Council of Turkey) under Project No. 224N535.

**Disclosure of Interests.** The authors have no competing interests to declare.

## References

1. Ali, H.M.U., Farooq, Q., Imran, A., El Hindi, K.: A systematic literature review on sentiment analysis techniques, challenges, and future trends. Knowledge and Information Systems **67**(5), 3967–4034 (2025). https://doi.org/10.1007/s10115-025-02365-x
2. Amendola, M., Cavaliere, D., De Maio, C., Fenza, G., Loia, V.: Towards echo chamber assessment by employing aspect-based sentiment analysis and gdm consensus metrics. Online social networks and media **39**, 100276 (2024)
3. Araci, D.T.: FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063 (2019), `https://arxiv.org/abs/1908.10063`
4. Aykaç, Y.E., Özkan, M., Samet, R.: Morpheme-aware hybrid sentiment analysis for azerbaijani: A lexicon–embedding fusion with domain adaptation. IEEE Access **14**, 18828–18856 (2026). https://doi.org/10.1109/ACCESS.2026.3659042
5. Aykaç, Y.E., Samet, R., Pashayev, A., Sabziev, E.: Sentiaznet: A polarity lexicon for azerbaijani. In: 2025 10th International Conference on Computer Science and Engineering (UBMK). pp. 1684–1689. IEEE (2025). https://doi.org/10.1109/UBMK67458.2025.11207041

6. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the association for computational linguistics **5**, 135–146 (2017)
7. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL). pp. 8440–8451. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.747, https://aclanthology.org/2020.acl-main.747/
8. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). pp. 4171–4186. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423/
9. D'aniello, G., Gaeta, M., La Rocca, I.: Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artificial Intelligence Review **55**(7), 5543–5574 (2022)
10. Guliyev, N., Rustamov, S., Rustamov, Z.: Analysis of public sentiment in azerbaijani news and social media. In: Proceedings of the 2024 International Conference on Computing, Machine Learning and Data Science (CMLDS '24). pp. 22:1–22:6. Association for Computing Machinery (2024). https://doi.org/10.1145/3661725.3661749
11. Jain, S., Wallace, B.C.: Attention is not explanation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers). pp. 3543–3556. Association for Computational Linguistics (2019), https://aclanthology.org/N19-1357/
12. Lamin, N.Z., Aziz, A.A.: Cross-lingual sentiment analysis in low-resource languages: A recent review on tasks, methods and challenges. International Journal of Advanced Computer Science & Applications **16**(11) (2025)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Loughran, T., McDonald, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance **66**(1), 35–65 (2011). https://doi.org/10.1111/j.1540-6261.2010.01625.x
15. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. vol. 30, pp. 4765–4774 (2017), https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
16. McNemar, Q.: Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika **12**(2), 153–157 (1947). https://doi.org/10.1007/BF02295996
17. Nasiopoulos, D.K., Roumeliotis, K.I., Sakas, D.P., Toudas, K., Reklitis, P.: Financial sentiment analysis and classification: A comparative study of fine-tuned deep learning models. International Journal of Financial Studies **13**(2), 75 (2025). https://doi.org/10.3390/ijfs13020075
18. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al.: Semeval-2016 task 5: Aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016). pp. 19–30 (2016)

19. Šmíd, J., Král, P.: Cross-lingual aspect-based sentiment analysis: A survey on tasks, approaches, and challenges. Information Fusion **120**, 103073 (2025). https://doi.org/10.1016/j.inffus.2025.103073
20. Šmíd, J., Přibáň, P., Král, P.: LACA: Improving cross-lingual aspect-based sentiment analysis with LLM data augmentation. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 839–853. Association for Computational Linguistics, Vienna, Austria (2025). https://doi.org/10.18653/v1/2025.acl-long.41, https://aclanthology.org/2025.acl-long.41/
21. Suhartono, D., Wongso, W., Tri Handoyo, A.: Idsarcasm: Benchmarking and evaluating language models for indonesian sarcasm detection. IEEE Access **12**, 87323–87332 (2024). https://doi.org/10.1109/ACCESS.2024.3416955
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30, pp. 5998–6008 (2017), https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
23. Wu, C., Lin, Z., Liu, Y., He, C., Yu, D., Yamada, I., Zhou, J., Miyao, Y.: M-ABSA: A multilingual dataset for aspect-based sentiment analysis. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. pp. 2525–2541. Association for Computational Linguistics, Miami, Florida, USA (Nov 2025). https://doi.org/10.18653/v1/2025.emnlp-main.128, https://aclanthology.org/2025.emnlp-main.128
24. Zhu, Z., Chen, H., Ye, X., Lyu, Q., Tan, C., Marasović, A., Wiegreffe, S.: Explanation in the era of large language models. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 5: Tutorial Abstracts). pp. 19–25 (2024)

# Section-Aware Retrieval for Turkish Case Law: A Case Study on Scanned Court Decisions

Kübra Kovayçin[1], Ece İrem Şişer[2], Derda Sina Günay[2], Yusuf Evren Aykaç[1,2], and Refik Samet[2]

[1] Ankara Yıldırım Beyazıt University, Ankara, Türkiye
[2] Ankara University, Ankara, Türkiye
yeaykac@aybu.edu.tr

**Abstract.** Finding relevant precedents often needs more than topical overlap: operative facts and judicial reasoning can be decisive. In Turkey, court decisions are often long and frequently shared as scanned (image-based) PDFs, making extraction and indexing harder. We present a section-aware retrieval case study on 1,000 Turkish court decisions about social media–related disputes. Using a vision–language model, each PDF is transcribed and organized into `Facts`, `Reasoning`, and `Verdict` segments; statute/article mentions are also extracted as lexical anchors. We compare a lexical baseline (BM25) with multilingual and legal bi-encoders adapted via parameter-efficient fine-tuning, and we test lexical–dense hybrid fusion, including a simple section-aware hybrid that scores sections separately under fixed token budgets. Because expert-labeled precedent links are not available, we use a self-alignment known-item diagnostic: an LLM generates a short, query-like summary, and the system tries to retrieve the decision it came from; full-document BM25 is included as a reference baseline. Under this proxy, BM25 over `Reasoning` is the strongest single-section index (R@1=0.744), and section-aware hybrid fusion with MPNet reaches R@1=0.780 and MRR@10=0.810, outperforming naive fusion while preserving section-level interpretability. This proxy setup is intended as a diagnostic rather than a full relevance benchmark; retrieval-time scoring is deterministic, using BM25 and cosine similarity over fixed embeddings.

**Keywords:** Legal information retrieval · precedent retrieval · section-aware retrieval · BM25 · neural bi-encoder models · sparse–dense hybrid fusion · large language models (LLMs) · Turkish case law

## 1 Introduction

Finding persuasive prior decisions is central to legal research and case preparation. In practice, a helpful precedent is not only about the same topic; it should also align with key facts and with how the court reasons about the law [7, 19].

Working with Turkish decisions adds practical issues. Turkish is morphologically rich, which can increase lexical mismatch in term-based retrieval [26]. Many

decisions are also distributed as scanned, image-based PDFs. When the source is noisy, straightforward OCR and rule-based section splitting can be brittle, and the resulting text may not be ready for retrieval without additional processing.

Evaluation is another limitation. Benchmarks such as COLIEE provide labeled links between cases [10], but we are not aware of a public Turkish dataset with expert-labeled precedent links for our target domain. For this reason, we use a controlled *self-alignment* proxy: for each decision we generate a short, query-like summary and ask the system to retrieve the same decision from the corpus. We treat this as a diagnostic for section usefulness and fusion choices, not as a substitute for expert relevance judgments.

In this work, we compile 1,000 Turkish judicial decisions involving social media–related disputes. We process the scanned PDFs with a vision–language model and obtain structured `Facts`, `Reasoning`, and `Verdict` segments (Section 3). We then compare sparse retrieval (BM25), dense bi-encoders, and simple lexical–dense hybrid fusion under identical section and token-budget constraints.

We study three *research questions*: (RQ1) Which sections are most useful under the self-alignment proxy? (RQ2) Does section-aware fusion help compared to naive concatenation under fixed token budgets? (RQ3) How do lexical anchors (e.g., statute mentions) and semantic similarity interact in early-rank retrieval?

*Contributions.* In this case study, we:

(1) **create** a section-structured corpus of 1,000 Turkish decisions by extracting role-labeled text and metadata from scanned PDFs with a VLM pipeline (with schema validation and reprocessing);
(2) **run** controlled comparisons of sparse, dense, and hybrid retrievers across different sections using the self-alignment proxy;
(3) **describe** a simple section-aware hybrid fusion template and report observations that may inform future, expert-labeled evaluations.

The remainder of the paper is organized as follows: Related Works reviews prior research; Method describes corpus construction, the proxy diagnostic, and retrieval models; Results reports quantitative findings; Discussion discusses implications and limitations; and Conclusion summarizes takeaways and future directions.

## 2 Related Work

We situate our study within prior work on legal case retrieval, long-document retrieval, Turkish legal NLP, and LLM-assisted preprocessing.

### 2.1 Legal Case Retrieval and Evaluation

Precedent retrieval is often framed as a *query-by-document* task, where both the query and the candidate are long, structured decisions and relevance depends on operative facts and legal reasoning [7, 19]. When expert labels exist,

benchmarks such as COLIEE enable more direct comparison [10]. Some work also uses citation or linkage structure with graph-based or transformer-based models [22,23], but this typically requires machine-readable links that are not available in many low-resource corpora. Because expert judgments are costly and jurisdiction-specific [7], proxy evaluations (e.g., self-retrieval or weak supervision) are often used for controlled diagnostics, but they do not replace true precedent relevance [24].

### 2.2 Sparse, Dense, and Hybrid Retrieval for Long Decisions

BM25 is still a common and efficient first-stage baseline in law, partly because legal text contains strong lexical anchors such as citations and terminology [18]. Dense bi-encoders can help with paraphrase and vocabulary mismatch [12], while learned-sparse and late-interaction models provide intermediate trade-offs [8,13]. Hybrid fusion methods (score fusion or rank fusion) combine lexical and semantic signals and can be more robust when scores are not well calibrated [5,28]. For long, heterogeneous legal decisions, several studies report benefits from passage- or structure-aware aggregation instead of naive full-text concatenation [1,2,14,20].

### 2.3 Domain Adaptation and Turkish Legal NLP

Domain-adapted encoders (e.g., Legal-BERT) and multilingual backbones such as XLM-R are commonly used starting points for legal retrieval [3,4]. Parameter-efficient fine-tuning (LoRA/QLoRA) can make adaptation feasible under limited compute [6,11]. For Turkish, resources remain more limited than for English, but Turkish legal encoders and prior-case retrieval studies have started to appear [16,27]. Public corpora and datasets are also emerging [15], although stable expert-labeled precedent links are still uncommon.

### 2.4 Large Language Models in the Pipeline

LLMs are increasingly used for summarization and extraction in modern IR pipelines [9]. In this paper, we restrict LLM usage to preprocessing steps (PDF transcription/structuring and query-surrogate summary generation). At retrieval time, we do *not* call an LLM; scoring is deterministic via BM25 and cosine similarity over fixed embeddings.

*Gap and motivation.* Much of the existing literature either assumes clean, machine-readable case text or relies on curated relevance labels/citation graphs. For Turkish case law in scanned PDF form, both assumptions can be fragile. Our goal is therefore modest: we provide a VLM-based preprocessing pipeline and a reproducible diagnostic that make it easier to compare section choices and simple fusion strategies before investing in expert labeling.

```
{
    "file_name": "2022_551.pdf",
    "meta_data": {
        "esas_no": "2021/1168",
        "karar_no": "2024/1919",
        "court_name": "İSTANBUL BÖLGE ADLİYE MAHKEMESİ"
    },
    "raw_text_preview": "T.C.\n\nİSTANBUL\n\nBÖLGE ADLİYE MAHKEMESİ\n\n12. HUKUK DAİRESİ\n\nDOSYA NO: 2022/551\n\nKARAR NO:
2024/1919\n\nTÜRKMİLLETİADINA\n\nİSTİNAFKARARI\n\nİNCELENEN KARARIN\n\nMAHKEMESİ: BAKIRKÖY 2. ASLİYE TİCARET MAHKEMESİ\n"
}
```


**Fig. 1.** Regex-based segmentation failure on a scanned decision.

## 3 Method

We next describe corpus construction, section extraction, retrieval models, and the proxy evaluation protocol. LLMs are used only for corpus construction and query-surrogate generation; retrieval itself uses only stored text, indices, and embeddings.

### 3.1 Data preprocessing

We convert image-based PDFs into structured, retrieval-ready text segments.

*Corpus.* We curated 1,000 Turkish judicial decisions in the domain of *Social Media Law*. Documents are primarily image-based PDFs with heterogeneous scan quality and layout artefacts (e.g., stamps, headers/footers, skew, and page-break fragmentation), which makes conventional OCR followed by rule-based section parsing brittle.

*Why regex segmentation is insufficient.* A simple baseline is OCR followed by regex patterns that try to split decisions into rhetorical roles. In practice, layout noise and court-specific templates often cause boundary shifts, missing headers, and merged sections, yielding incomplete or unstructured outputs. Figure 1 shows one typical failure case.

In such cases, the pipeline may not reliably separate `Facts`, `Reasoning`, and `Verdict`, which can introduce noise into downstream indices.

*VLM-based Multimodal Extraction and Refinement.* For our multimodal extraction pipeline, we utilized `Amazon Nova-2-Lite-V1`, selected for its cost-efficiency—specifically its integrated visual processing without additional surcharges—and its competitive performance with a 1M token context window, comparable to Gemini-class models. The pipeline transforms legal document images into structured text, segmenting each decision into three core components (`Facts`, `Reasoning`, and `Verdict`) and generating dual-purpose summaries (Figure 2). To ensure data integrity, we implemented custom validation algorithms to detect and discard records that were incomplete or truncated due to token limits. For these cases, we utilized a refined inference configuration with `max_tokens` set to 3,000 and a `repetition_penalty` of 1.2 during the re-processing stage. Furthermore, to mitigate "infinite loop" issues observed during legal statute extraction, system prompts were strictly constrained with instructions to mention each law article only once.

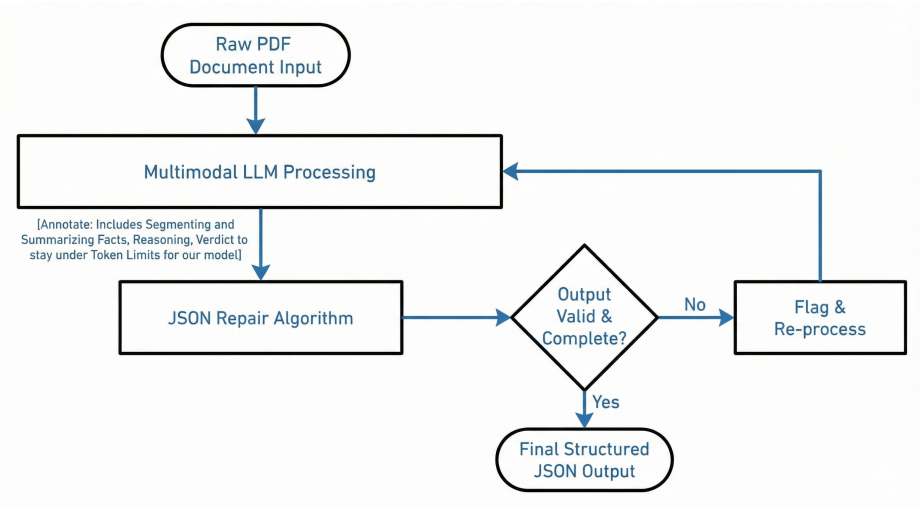



**Fig. 2.** Preprocessing pipeline for scanned PDFs.


```
{
    "meta_data": {
        "file_name": "2019_1462.pdf",
        "court_name": "İstanbul Bölge Adliye Mahkemesi 16. Hukuk Dairesi",
        "esas_no": "2019/1462",
        "karar_no": "2019/1457",
        "case_subject": "İhtiyati Tedbir",
        "dava_tarihi": "06/03/2019",
        "karar_tarihi": "05/07/2019"
    },
    "rrl_segments": {
        "facts_text": "İddialar, vekilin müvekkilinin ticari markalarını kullanmak için internet sitelerinde reklam ve tanıtım faaliyetleri yapmasıdır. Mahkeme, istinaf dilekçesi ve ek
tedbir taleplerini değerlendirmiştir.",
        "reasoning_text": "Mahkeme, Hükim 353 ve 356 maddeleri çerçevesinde, istinaf ve tedbir taleplerini reddetmiştir. İstinaf istemi reddedilmiş ve dava süreci devam etmiştir.",
        "verdict_text": "Dair, dosya üzerinde yapılan inceleme sonucu 05/07/2019 tarihinde oy birliği ile kesin olarak karar verildi."
    },
    "structural_features": {
        "mentioned_laws": [
            "Hukuk 353",
            "Hukuk 356",
            "Hukuk 353/1-b-1"
        ]
    },
    "summary_for_human": "Mahkeme, vekilin istinaf ve tedbir taleplerini reddetmiştir. İstinaf istemi reddedilmiş ve dava süreci devam etmiştir.",
    "summary_for_model": "Dava, ticari marka ihlali iddiası üzerine açılmıştır. Mahkeme, istinaf ve tedbir taleplerini reddetmiş ve dava süreci devam etmiştir."
}
```


**Fig. 3.** Example structured JSON record used for indexing and evaluation.

*Structured JSON Output.* The resulting dataset is stored in a structured JSON format, where each record comprises four key modules: (i) **Metadata**, containing case-specific details such as file name, court, case and decision dates, docket numbers, and subject matter; (ii) **Rhetorical Role Labeling (RRL)**, consisting of the role-labeled text for `Facts`, `Reasoning`, and `Verdict`; (iii) **Structural Features**, capturing extracted statutes (e.g., TCK, FSEK) used as lexical anchors; and (iv) **Summaries**, including a `summary_for_human` for general readability and a `summary_for_model` designed for self-retrieval and proxy evaluation tasks (Figure 3).

*Summary generation and overlap.* `summary_for_model` is generated using an LLM from the extracted content of the same decision and is used only as a query surrogate for the self-alignment diagnostic (Section 3.2). Because the query is derived from the target decision, lexical overlap is expected; we therefore treat

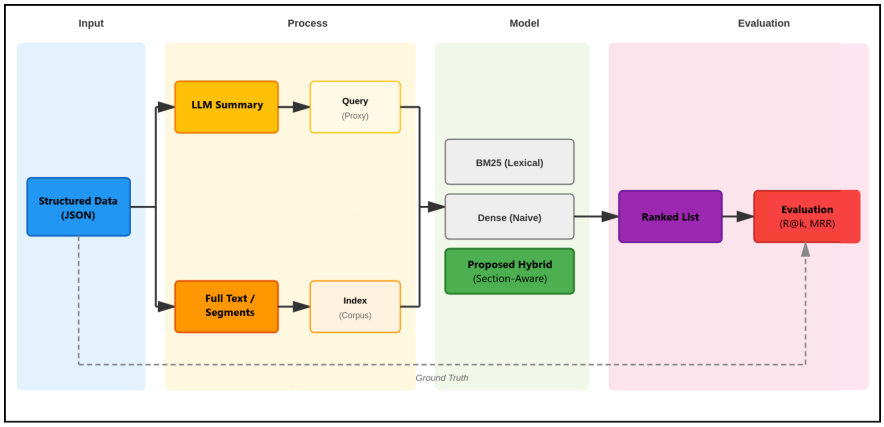



**Fig. 4.** Self-alignment diagnostic used for section-aware retrieval evaluation.

full-document BM25 as a reference baseline for this proxy and focus on relative section effects and fusion behavior.

*Validation and reprocessing.* We validate schema completeness and re-run extraction for invalid outputs. About 3% of cases required reprocessing due to segmentation shifts in complex layouts. To reduce repetitive generations, we use deterministic decoding (`temperature=0`) with a mild repetition penalty.

*Extraction quality and downstream impact.* This relatively small reprocessing rate should not be interpreted as evidence that extraction noise has no effect on retrieval performance. In the current study, we did not separately quantify the downstream impact of such errors, since the issue was identified earlier during segment extraction and completeness checking, before model training. We therefore treat remaining extraction imperfections as part of the practical preprocessing noise inherent to scanned legal documents. A more direct analysis of how extraction quality influences downstream retrieval performance is left for future work.

### 3.2 Section-aware self-alignment diagnostic

Because no expert-labeled Turkish precedent benchmark is available for our target domain, we use a controlled known-item proxy.

No Turkish benchmark provides expert-labeled precedent links for our target domain. We therefore use a reproducible *self-alignment* known-item proxy to compare section choices and lexical–dense fusion under identical noise and length constraints (Figure 4). This proxy is best read as a diagnostic rather than true precedent relevance; we include full-document BM25 as a reference baseline.

*Proxy task.* Let $\mathcal{D} = \{d_i\}_{i=1}^{N}$ denote the corpus of $N = 1{,}000$ decisions. For each $d_i$, we query with $q_i = \texttt{summary_for_model}(d_i)$ and treat $d_i$ as the single relevant target. A retriever produces a ranked list over $\mathcal{D}$; success is measured by the rank of $d_i$.

*Metrics.* With one relevant target per query, we report Recall@1/5/10 and MRR@10.

### 3.3 Retrieval models

We evaluate three families of first-stage retrievers: sparse (lexical), dense (semantic), and simple hybrids. None of these models call an LLM at retrieval time.

*BM25 (sparse).* We use BM25 [18] as the lexical baseline. To isolate section effects (RQ1), we build separate indices over `facts`, `laws`, `facts+laws`, `facts+laws+subject`, `facts+laws+subject+verdict`, `reasoning`, and `full`.

*Dense bi-encoders.* Dense retrievers encode queries and candidates into a shared vector space and rank by cosine similarity [17]. We evaluate multilingual encoders (MiniLM, MPNet) [21,25] and legal/multilingual backbones (BERTurk-Legal, XLM-R) [4,16]. For XLM-R, we compare full fine-tuning and parameter-efficient variants (LoRA and 4-bit QLoRA) [6,11]. Weak supervision uses $q = \texttt{summary_for_model}(d)$ with the positive segment $d^+ = \texttt{reasoning_text}(d)$. Hard negatives are mined with BM25 within the training split to encourage discrimination among lexically similar decisions.

*Hybrid fusion.* We combine lexical and dense signals via late fusion:

$$s_{\text{hyb}}(q,d) = \alpha \cdot \tilde{s}_{\text{BM25}}(q,d) + (1-\alpha) \cdot \tilde{s}_{\text{dense}}(q,d), \tag{1}$$

with per-query min–max normalization

$$\tilde{s}(q,d) = \frac{s(q,d) - \min_{d'} s(q,d')}{\max_{d'} s(q,d') - \min_{d'} s(q,d') + \epsilon}. \tag{2}$$

To reduce semantic dilution under fixed token budgets (RQ2) and to combine complementary evidence (RQ3), we use a simple section-aware hybrid that fuses BM25 over `facts+laws` and `reasoning` with dense similarity over `reasoning`:

$$s_{\text{sec-hyb}}(q,d) = \beta \cdot \tilde{s}_{\text{BM25}}^{(\text{facts+laws})}(q,d) + \gamma \cdot \tilde{s}_{\text{BM25}}^{(\text{reason})}(q,d) + (1-\beta-\gamma) \cdot \tilde{s}_{\text{dense}}^{(\text{reason})}(q,d), \tag{3}$$

where $\beta, \gamma \geq 0$ and $\beta + \gamma \leq 1$. As an alternative that avoids score calibration, we also evaluate Reciprocal Rank Fusion (RRF) [5]:

$$\text{RRF}(d) = \sum_{m \in \mathcal{M}} \frac{1}{k + \text{rank}_m(d)}. \tag{4}$$

Cross-encoder reranking is included only as an ablation; in our runs, gains were small under our length constraints, so we do not use it in the main pipeline.

### 3.4 Experimental settings

We summarize the main experimental choices below for reproducibility.

*Splits.* We use the full corpus of $N = 1{,}000$ structured decisions and a fixed 70/10/20 train/validation/test split (700/100/200) with decision-level separation. Dense training (including BM25-based negative mining) uses only the training split; hyperparameter and fusion-weight selection uses validation; we report metrics on the held-out test set. To control leakage, test decisions are excluded from training both as positives and as mined negatives.

*BM25 preprocessing.* We build BM25 indices for each segment configuration. Text processing includes Unicode and whitespace normalization, lowercasing, and lightweight stopword filtering tailored to Turkish and common legal boilerplate. To improve citation matching, statutory references (e.g., "TTK 55/1-a-1") are normalized into single tokens (e.g., "ttk_55_1_a_1") while preserving digits. For selected segment configurations, contextual fields (e.g., mentioned laws or subject) may be upweighted via token repetition. Unless otherwise stated, BM25 uses $k_1 = 1.2$ and $b = 0.75$.

*Dense training.* Dense retrievers are trained as bi-encoders with cosine similarity (dot product of L2-normalized embeddings). We fine-tune using TripletLoss (margin 0.2). For each query, a hard negative is sampled from BM25's top-50 results (excluding the positive), drawn from ranks 6–30 to avoid trivial negatives. Inputs are truncated to a maximum sequence length of 256 tokens. We run an epoch sweep over $\{1, \ldots, 7\}$ with fixed random seed 42, batch size 8, learning rate $2 \cdot 10^{-5}$, and mixed-precision training (AMP). The best checkpoint is selected based on validation MRR@10 of the section-aware hybrid system.

*Fusion and reproducibility.* Score-level fusion uses per-query min–max normalization (Eq. 2). Unless otherwise stated, we use $\alpha = 0.5$ for the basic hybrid (Eq. 1) and $(\beta, \gamma) = (0.45, 0.20)$ for the section-aware hybrid (Eq. 3). These weights were selected on the validation split of the current dataset and should therefore be interpreted as task-specific settings rather than generally optimal values. We intentionally kept the fusion design simple and interpretable, since the goal of this study is not to propose a heavily tuned or learned fusion mechanism, but to examine section utility and hybrid behavior in a controlled, low-resource diagnostic setting. RRF uses $k = 60$ (Eq. 4) and does not require score calibration. LLM usage is limited to corpus construction and summary generation; retrieval-time scoring is deterministic given fixed indices and embeddings.

## 4 Results

We report the main quantitative findings under the self-alignment proxy below. As noted above, these results are best read as a diagnostic for recovering the source decision, not as expert-labeled precedent relevance.

**Table 1.** BM25 across indexed segments under the self-alignment proxy (query=`summary_for_model`). Full-document BM25 is included as a reference baseline for this proxy.

| BM25 Index Segment | R@1 | R@5 | R@10 |
|---|---|---|---|
| Facts | 0.542 | 0.687 | 0.715 |
| Laws | 0.111 | 0.192 | 0.236 |
| Facts + Laws | 0.629 | 0.757 | 0.799 |
| Facts + Laws + Subject | 0.645 | 0.778 | 0.813 |
| Facts + Laws + Subject + Verdict | 0.728 | 0.851 | 0.894 |
| Reasoning | 0.744 | 0.833 | 0.865 |
| Full Document | **0.913** | **0.963** | **0.974** |

*BM25 segment effects (RQ1).* Table 1 reports BM25 across section-wise indices under the self-alignment proxy. We include full-document BM25 as a reference baseline for this proxy (R@10=0.974). Among single sections, `Reasoning` gives the highest scores in our setup (R@1=0.744, R@10=0.865). Indexing `Facts + Laws + Subject + Verdict` also helps at early ranks (R@1=0.728, R@10=0.894), bringing performance closer to the full-document reference. In contrast, `Laws`-only indexing is weak, and `Facts`-only remains below `Reasoning`. Overall, this suggests that much of the signal for this proxy task comes from legal reasoning, while contextual fields (subject and verdict) can provide additional lexical cues.

*Dense vs. hybrid retrieval (RQ2–RQ3).* In this weakly supervised known-item setting, dense-only retrievers are generally weaker than BM25, suggesting that lexical overlap and statutory cues remain highly informative under this proxy. However, combining dense retrieval with BM25 consistently improves performance across models. Table 2 shows representative backbones and indicates that section-aware fusion usually yields the strongest early-rank results. In particular, MPNet with section-aware fusion achieves the best overall R@1 and MRR@10 scores (R@1=0.780, MRR@10=0.810), while the highest R@5 score is 0.870, reached by both XLM-R (FULL) and XLM-R (LoRA) with section-aware fusion, and XLM-R (LoRA) attains the highest R@10 (0.900). Overall, these results suggest that reasoning-focused semantic matching is most effective when combined with explicit lexical anchoring, rather than compressed into a single naive concatenated representation. This brings hybrid performance closer to the full-document BM25 reference while preserving section-level scoring and interpretability.

*Training segment pairs.* When training dense encoders, using `summary_for_model` as queries and `Reasoning` as positives gave the best dense-only results in our runs. Training on full-text or multi-section concatenations tended to underperform, which is consistent with signal dilution under fixed token budgets (table omitted for space).

**Table 2.** Dense vs. hybrid retrieval results on the test split (seed 42). Best epochs are selected based on validation MRR@10 of the section-aware hybrid system.

| **Model / System** | **R@1** | **R@5** | **R@10** | **MRR@10** |
|---|---|---|---|---|
| MiniLM Dense | 0.390 | 0.560 | 0.620 | 0.458 |
| MiniLM Hybrid (naive) | 0.640 | 0.780 | 0.850 | 0.707 |
| MiniLM Hybrid (section-aware) | 0.730 | 0.860 | 0.880 | 0.788 |
| MPNet Dense | 0.490 | 0.620 | 0.650 | 0.544 |
| MPNet Hybrid (naive) | 0.670 | 0.800 | 0.860 | 0.728 |
| MPNet Hybrid (section-aware) | **0.780** | 0.860 | 0.890 | **0.810** |
| BERTurk-Legal Dense | 0.430 | 0.520 | 0.590 | 0.468 |
| BERTurk-Legal Hybrid (naive) | 0.640 | 0.830 | 0.850 | 0.717 |
| BERTurk-Legal Hybrid (section-aware) | 0.710 | 0.860 | 0.880 | 0.775 |
| XLM-R (FULL) Dense | 0.380 | 0.460 | 0.460 | 0.409 |
| XLM-R (FULL) Hybrid (naive) | 0.620 | 0.770 | 0.820 | 0.683 |
| XLM-R (FULL) Hybrid (section-aware) | 0.690 | **0.870** | 0.890 | 0.765 |
| XLM-R (LoRA) Dense | 0.380 | 0.460 | 0.470 | 0.413 |
| XLM-R (LoRA) Hybrid (naive) | 0.620 | 0.770 | 0.820 | 0.683 |
| XLM-R (LoRA) Hybrid (section-aware) | 0.700 | **0.870** | **0.900** | 0.772 |
| XLM-R (QLoRA, 4bit) Dense | 0.415 | 0.470 | 0.510 | 0.442 |
| XLM-R (QLoRA, 4bit) Hybrid (naive) | 0.620 | 0.765 | 0.810 | 0.682 |
| XLM-R (QLoRA, 4bit) Hybrid (section-aware) | 0.745 | 0.860 | 0.895 | 0.796 |

## 5 Discussion

We interpret the proxy results below and highlight their main limitations.

Under the self-alignment proxy, section choice has a clear effect. `Reasoning` tends to be more informative than `Facts` or `Laws` alone, and dense-only models trained with weak supervision do not match lexical BM25 in this setup. However, simple late fusion of lexical and dense signals often improves early ranks, especially when dense similarity focuses on `Reasoning`.

*Implications for section choice.* Our emphasis on the `Reasoning` section was not chosen arbitrarily. It was also informed by discussions with legal practitioners, who indicated that the reasoning part of a decision is often the most critical component when assessing precedent value. In this sense, our findings do not merely suggest a data-driven pattern, but also provide empirical support for an intuition already present in legal practice. Under our proxy setting, the experimental results reinforce the view that `Reasoning` carries a particularly strong retrieval signal compared with `Facts` or isolated law mentions alone.

*Why naive concatenation can hurt.* Concatenating heterogeneous sections into a single encoder input can underperform section-wise scoring. Long inputs trigger truncation, boilerplate templates may dominate the representation, and frequent citations can introduce false positives. Section-aware fusion can reduce these issues by anchoring with BM25 over `facts+laws` and focusing dense similarity on `reasoning`.

*Limitations of the proxy.* The proxy does not measure true precedent relevance: queries are summary surrogates derived from the same decision, and full-document BM25 is naturally advantaged (reported here as a reference baseline). Results may vary with summary style, OCR/segmentation noise, and the small training regime ($N$=1,000; 700 train), which can limit dense adaptation under domain/style shift. In addition, because the query is derived from the target decision itself, the setup should be interpreted as a controlled known-item diagnostic with expected lexical overlap rather than a realistic substitute for expert-judged precedent search. A necessary next step is expert pooling with graded relevance and standard ranked metrics [16].

## 6 Conclusion

In this case study, we explored section-aware retrieval for Turkish legal decisions using a reproducible self-alignment proxy. Within this setup, we observed that the `Reasoning` section often carries more retrieval signal than `Facts` or `Laws` alone, and that simple lexical–dense fusion can help under fixed token budgets. We emphasize that our evaluation is a diagnostic rather than expert-labeled precedent retrieval. The corpus was intentionally limited to social-media-related disputes so that section effects and fusion behavior could be studied in a more controlled setting, with less heterogeneity in legal terminology, case structure, and retrieval signals. While this limits direct generalization to other legal domains, it also makes the present study methodologically cleaner as an initial domain-specific diagnostic. Future work will add expert-labeled evaluation with pooling and graded relevance (and/or practitioner user studies), extend the corpus gradually to additional legal domains, explore paragraph-level aggregation baselines, and revisit cross-encoder reranking under tighter evidence selection.

## References

1. Althammer, S., Hofstätter, S., Sertkan, M., Verberne, S., Hanbury, A.: Parm: A paragraph aggregation retrieval model for dense document-to-document retrieval. In: European Conference on Information Retrieval. pp. 19–34. Springer (2022)
2. Askari, A., Verberne, S., Abolghasemi, A., Kraaij, W., Pasi, G.: Retrieval for extremely long queries and documents with rprs: a highly efficient and effective transformer-based re-ranker. ACM Transactions on Information Systems **42**(5), 1–32 (2024)

3. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I.: Legal-bert: The muppets straight out of law school. arXiv preprint arXiv:2010.02559 (2020)
4. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th annual meeting of the association for computational linguistics. pp. 8440–8451 (2020)
5. Cormack, G.V., Clarke, C.L., Buettcher, S.: Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. pp. 758–759 (2009)
6. Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L.: Qlora: Efficient finetuning of quantized llms. Advances in neural information processing systems **36**, 10088–10115 (2023)
7. Feng, Y., Li, C., Ng, V.: Legal case retrieval: A survey of the state of the art. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 6472–6485 (2024)
8. Formal, T., Piwowarski, B., Clinchant, S.: Splade: Sparse lexical and expansion model for first stage ranking. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2288–2292 (2021)
9. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., Wang, H.: Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997 **2**(1) (2023)
10. Goebel, R., Kano, Y., Kim, M.Y., Rabelo, J., Satoh, K., Yoshioka, M.: Overview and discussion of the competition on legal information, extraction/entailment (coliee) 2023. The Review of Socionetwork Strategies **18**(1), 27–47 (2024)
11. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
12. Karpukhin, V., Oguz, B., Min, S., Lewis, P.S., Wu, L., Edunov, S., Chen, D., Yih, W.t.: Dense passage retrieval for open-domain question answering. In: EMNLP (1). pp. 6769–6781 (2020)
13. Khattab, O., Zaharia, M.: Colbert: Efficient and effective passage search via contextualized late interaction over bert. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. pp. 39–48 (2020)
14. Nigam, S.K., Dubey, T., Shallum, N., Bhattacharya, A.: Segment first, retrieve better: Realistic legal search via rhetorical role-based queries. arXiv preprint arXiv:2508.00679 (2025). https://doi.org/10.48550/arXiv.2508.00679, https://arxiv.org/abs/2508.00679
15. Öztürk, C.E., Köksal, Ö., Özçelik, Ş.B., Koç, A.: Prior case retrieval for the court of cassation of turkey. In: 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). pp. 539–544. IEEE (2022)
16. Öztürk, C.E., ÖzçelıK, Ş.B., Koç, A.: A transformer-based prior legal case retrieval method. In: 2023 31st Signal Processing and Communications Applications Conference (SIU). pp. 1–4. IEEE (2023)
17. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
18. Robertson, S., Zaragoza, H., et al.: The probabilistic relevance framework: Bm25 and beyond. Foundations and Trends® in Information Retrieval **3**(4), 333–389 (2009)

19. Sansone, C., Sperlí, G.: Legal information retrieval systems: state-of-the-art and open issues. Information Systems **106**, 101967 (2022)
20. Shao, Y., Mao, J., Liu, Y., Ma, W., Satoh, K., Zhang, M., Ma, S.: BERT-PLI: Modeling paragraph-level interactions for legal case retrieval. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). pp. 3501–3507 (2020). https://doi.org/10.24963/ijcai.2020/484, https://www.ijcai.org/proceedings/2020/484
21. Song, K., Tan, X., Qin, T., Lu, J., Liu, T.Y.: Mpnet: Masked and permuted pre-training for language understanding. Advances in neural information processing systems **33**, 16857–16867 (2020)
22. Su, W., Ai, Q., Wu, Y., Ma, Y., Li, H., Liu, Y., Wu, Z., Zhang, M.: Caseformer: Pre-training for legal case retrieval based on inter-case distinctions. arXiv preprint arXiv:2311.00333 (2023). https://doi.org/10.48550/arXiv.2311.00333, https://arxiv.org/abs/2311.00333
23. Tang, Y., Qiu, R., Yin, H., Li, X., Huang, Z.: CaseLink: Inductive graph learning for legal case retrieval. arXiv preprint arXiv:2403.17780 (2024). https://doi.org/10.48550/arXiv.2403.17780, https://arxiv.org/abs/2403.17780
24. Voorhees, E.M.: Report on trec-9. In: ACM SIGIR Forum. vol. 34, pp. 1–8. ACM New York, NY, USA (2000)
25. Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., Zhou, M.: Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Advances in neural information processing systems **33**, 5776–5788 (2020)
26. Yılmaz, K.E., Arslan, A., Yilmazel, O.: Turkish text retrieval experiments using Lemur toolkit. arXiv preprint arXiv:1405.1740 (2014), https://arxiv.org/abs/1405.1740
27. Zeidi, F., Amasyali, M.F., Erol, Ç.: Legalturk optimized bert for multi-label text classification and ner. arXiv preprint arXiv:2407.00648 (2024)
28. Zhou, Y., Huang, H., Wu, Z.: Boosting legal case retrieval by query content selection with large language models. In: Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region. pp. 176–184 (2023)

# A Fuzzy Neural Network Approach for Intelligent Environmental Monitoring and Anomaly Detection in IoT Networks

Majidov Anvarkhon[1][0009-0009-3797-1533]

[1] National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan
anvarxonmajidov@gmail.com

**Abstract.** The rapid deployment of Internet of Things (IoT) technologies has significantly expanded the capabilities of large-scale environmental monitoring systems. How-ever, the heterogeneity of sensor devices and the openness of communication channels increase the risk of anomalous data transmission and malicious interference. This paper proposes an intelligent environmental monitoring approach based on a fuzzy neural network for anomaly detection in IoT networks. The proposed method combines fuzzy logic and a multilayer neural network to analyze sensor data packets using nine network traffic features. Adaptive adjustment of membership functions during neural network training enables probabilistic evaluation of normal and anomalous behavior. The approach is integrated into a software-defined networking (SDN) architecture, allowing real-time traffic analysis, node validation, and dynamic synthesis of packet filtering rules. Simulation results demonstrate high anomaly detection accuracy with moderate processing overhead, confirming the effectiveness of the proposed solution for secure and scalable environmental monitoring.

**Keywords:** Environmental monitoring, Internet of Things, fuzzy logic, neural networks, anomaly detection, software-defined networking.

## 1 Inroduction

The rapid adoption of IoT technologies in various domains has significantly increased machine-to-machine interactions, leading to the collection and exchange of large volumes of data from spatially distributed sensors and devices. While this has enabled intelligent environmental monitoring and improved operational efficiency, it has also introduced new challenges in information security. The growing number of IoT devices, coupled with their connectivity over distributed telecommunication channels, has escalated the risk of cyber-attacks and anomalous behavior in IoT networks [1, 2]. Among the most critical threats in such networks are flooding and Distributed Denial of Service (DDoS) attacks, which generate excessive requests to network nodes, potentially disrupting the normal operation of critical Environmental Monitoring systems [3].

To address these challenges, intelligent Environmental Monitoring systems based on fuzzy logic and neural networks have been proposed. These systems are capable of analyzing, recognizing, and classifying network traffic patterns to detect anomalies in real time. In critical infrastructures, such as road transport systems, environmental monitoring involves the collection, processing, and storage of data from multiple sources, including photo radar complexes, video surveillance cameras, and other IoT sensors [4, 5]. Such data provides information about traffic conditions, accidents, and other environmental events, which is essential for ensuring public safety under concepts like “Safe City” and “Safe Roads.”

Recent advances in SDN have introduced new possibilities for intelligent Environmental Monitoring and anomaly detection in IoT networks [6, 7]. SDN provides centralized control of network switches, separating the logical management layer from the data forwarding layer, which facilitates dynamic reconfiguration, traffic analysis, and threat mitigation [8–10]. However, the centralized nature of SDN introduces limitations, particularly in terms of scalability and responsiveness during distributed attacks. In the case of large-scale DDoS attacks, the SDN controller may struggle to process incoming requests and update flow table rules in a timely manner [11, 12]. To overcome these limitations, SDN architectures are increasingly integrated with machine learning, fuzzy logic, neural networks, and statistical analysis techniques for proactive detection and mitigation of network anomalies [2, 13].

Several studies have explored combining SDN with advanced computational methods for IoT security. For example, platforms have been proposed for real-time detection of DDoS attacks using SDN in corporate networks [14], while others integrate SDN, blockchain, and network function virtualization to enhance IoT network security [15]. Such approaches address critical challenges, including low detection efficiency, lack of standardization, and high latency during large-scale attacks. Additionally, SDN-based solutions for low-power sensor networks enable real-time anomaly detection with minimal computational overhead [16]. Other works have explored Environmental Monitoring flooding attacks in SDN controllers’ data caches [17] and securing heterogeneous networks using blockchain-based distributed ledgers for data integrity and authentication [18].

Building on these insights, this article presents a secure and intelligent IoT Environmental Monitoring system that integrates SDN technology, fuzzy logic, and a multilayer neural network for real-time environmental monitoring and anomaly detection. The proposed system leverages distributed ledger technology for secure data storage, smart contracts for validating network nodes, synthesis of flow table rules, and comprehensive traffic analysis and filtering. This approach enables robust detection of network anomalies, mitigates cyber threats, and ensures reliable operation of critical IoT Environmental Monitoring infrastructures.

## 2 Materials and methods

The main function of the subsystem for protecting operations on collecting and transmitting network traffic from distributed nodes in the Environmental Monitoring

system is to analyze data packets from authorized nodes in order to detect atypical packets with possible malicious content. The architecture of the developed traffic analysis subsystem includes a parser, an analyzer, a clustering module with a neural network, and a logging module (Fig. 1).

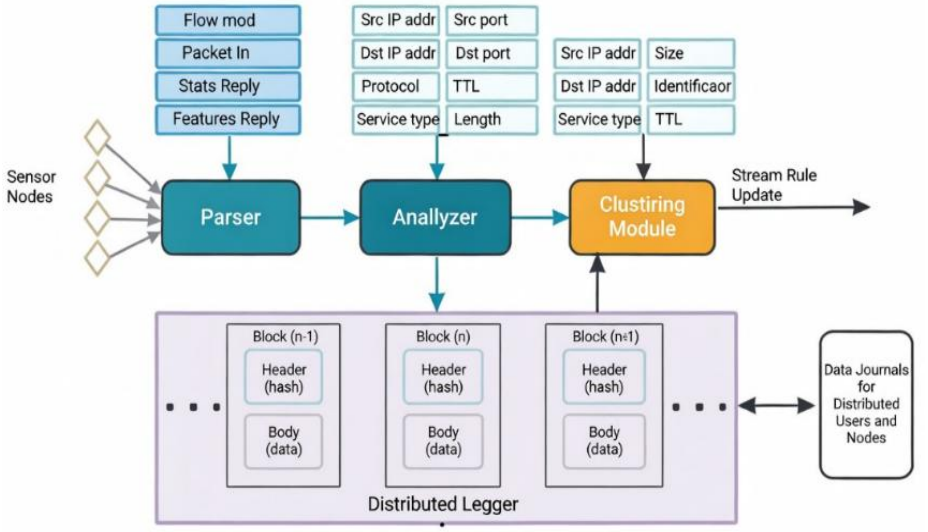



Fig. 1. Subsystem for analysis and detection of atypical data packets

Parsing and intelligent analysis modules are required to detect data packets with characteristics different from those usually transmitted by authorized nodes. The analysis procedure is implemented in the SDN network controller and allows identifying atypical behavior of sensor nodes and user mobile devices. Packets are selected by the parser and transmitted to the analyzer. The analysis module extracts packet characteristics from the headers of protocol data units. The characteristics are used to cluster data units in the feature space in order to determine the degree of similarity with the centroids of clusters of typical and atypical packets formed during the training of the neural network. Clustering allows detecting facts of modification of data packet headers and compromise of sensor nodes. For sets of characteristics of atypical packets, the controller creates filtering rules in flow tables, and for normal packets – retransmission rules.

The packet clustering method is implemented in a combined manner using fuzzy logic and a neural network. An example of such an approach is the intrusion detection technology in the ANFIS (Adaptive Network-based Fuzzy Inference System) [19]. To synthesize the feature space and perform clustering, it is proposed to use nine characteristics extracted from the headers of IP packets and TCP segments, namely the

source IP address (IP_src), destination IP address (IP_Dst), header length (Head_Lng), total packet length (Total_Lng), fragment identifier (Frag_Id), packet lifetime (TTL), service type (Srv_type), sender (Port_Srv) and recipient (Port_Dst) protocol ports. The initial training of the neural network is performed based on the sets of characteristics of packets from unauthorized nodes, and the identified sets of characteristics of atypical packets are subsequently used for training. The parameters of the function of packet membership in clusters are configured using a neural network training algorithm, and the conclusion about membership and the degree of similarity with cluster centroids is formed using a fuzzy logic apparatus. The fuzzy clustering method is based on the fuzzy c-means algorithm [20] and includes the following steps:

1. The number of packet clusters $M$ is set, which is then adjusted during the training process, the degree of fuzziness of the objective function $m > 1$ is selected.

2. The input sets of packet characteristics are feature vectors $X_j (j = 1, \dots, N)$. The vector defines a point in space that can belong to clusters with centroids $C^{(k)} (k = 1, \dots, M)$ and a probability membership function ${\mu_{X_j}}^{(k)}$, where $0 < {\mu_{X_j}}^{(k)} < 1$, $\sum_{k=1}^{M} {\mu_{X_j}}^{(k)} = 1$, which acts as the degree of similarity to the centroid and is defined as the distance ${D_{X_j}}^{(k)}$.

3. The points are randomly distributed among the clusters. The distribution of points is determined by the matrix of degrees of similarity with centroids in the feature space with nine coordinates $(x_{i1}, x_{i2} x_{i3} x_{i4} x_{i5} x_{i6} x_{i7} x_{i8} x_{i9})$.

4. The coordinates of the centroids of clusters $C^{(k)} (k = 1, \dots, M)$ are calculated by calculating the average proximity of cluster points:

$$C_k = \frac{\sum_{k=1}^{M} ({\mu_{X_j}}^{(k)} X_j)}{\sum_{k=1}^{M} ({\mu_{X_j}}^{(k)})}. \quad (1)$$

5. The distances between the points and the cluster centroids are determined:

$${D_{X_j}}^{(k)} = \sqrt{\| X_j - C^{(k)} \|^2}. \quad (2)$$

6. The degrees of membership of the points in the clusters are recalculated and the distribution matrix of the points is updated:

$$\mu_{X_j}{}^{(k)} = \frac{1}{\left(\sum_{k=1}^{M} \frac{D_{X_i}{}^{(k)}}{D_{X_j}{}^{(k)}}\right)^{2/m-1}}), \quad (3)$$

where $m > 1$ is the clustering fuzziness coefficient.

7. To stop the iteration process, the parameter $\varepsilon > 0$ is set. If the condition $\{|\mu_{X_j}{}^{(k)} - \mu_{X_j}{}^{(k-1)}|\} < \varepsilon$ is not met, then the transition to point 5 is performed.

The clustering method training procedure is implemented using a neural network, which is a five-layer structure without feedback with weight coefficients and activation functions (Fig. 2). The adaptive Takagi-Sugeno-Kang (TSK) model is chosen as a basis [21]. The output signal is determined by the aggregation function for $M$ rules and $N$ variables (in our case, there are $N = 9$ packet characteristics at the network input):

$$y(x) = \frac{\sum_{k=1}^{M}(\mathcal{W}_k * y_k(x))}{\sum_{k=1}^{M}(\mathcal{W}_k)}, \quad (4)$$

where $y_k(x) = z_{k0}\sum_{j=1}^{N}(z_{kj}x_j)$is the $i$-th polynomial component of the approximation, weights ($\mathcal{W}_i$ represent the degree of fulfillment of the conditions of the rule ($\mathcal{W}_k = \mu_A{}^{(k)}(x_j)$.

The membership function or fuzzification $\mu_A{}^{(k)}$ for the variable $x_j$ is represented by the Gaussian function

$$(\mathcal{W}_k = \mu_A{}^{(k)}(x_j) = \prod_{j=1}^{N}\left[\frac{1}{1+\left(\frac{\left(x_j - c_j{}^{(k)}\right)}{\sigma_j{}^{(k)}}\right)^{2b_j{}^{(k)}}}\right], \quad (5)$$

where $k$ - is the number of membership functions ($k = 1 \ldots M$); $j$ - is the number of variables ($N = 9$); $c_j{}^{(k)}, \sigma_j{}^{(k)}, b_j{}^{(k)}$ - parameters of the Gaussian function that determine its center, width and shape of the $k$-th membership function of the $j$-th variable.

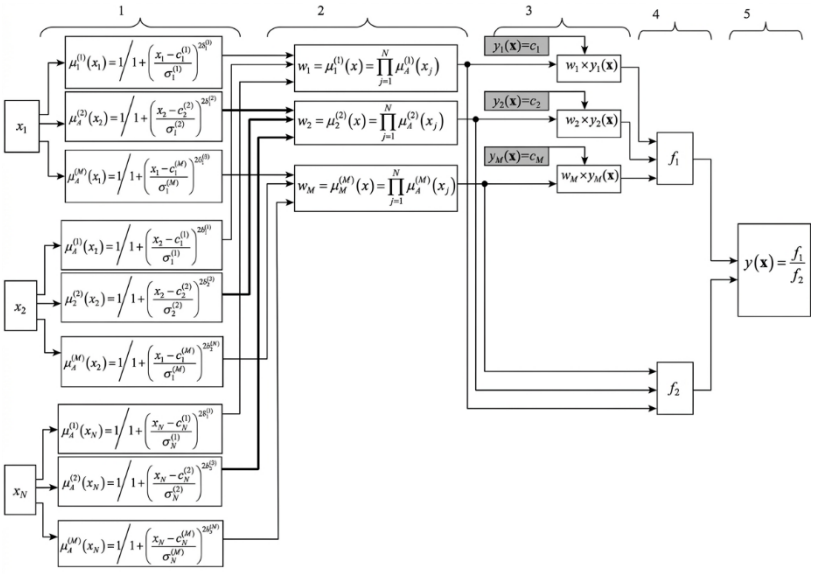


**Fig. 2.** Architecture of fuzzy neural network for clustering network traffic packets

The rules for deriving output variables $Y = (y_1, y_2, \dots, y_M)$ for a set of variables $X = (x_1, x_2, \dots, x_9)$, taking on a set of values $A_j{}^{(k)}$, represent a matrix of membership function values of size $9 \times M$:

$R1$: if $x_1^{(1)} \in A_1^{(1)}$ and $x_2^{(1)} \in A_2^{(1)}$and, …, and $x_9^{(1)} \in A_9^{(1)}$, then $y_1(x) = z_{10} \sum_{j=1}^{9} (z_{1j} x_j)$, … … … … (6)

$R_M$: if $x_1^{(M)} \in A_1^{(M)}$ and $x_2^{(M)} \in A_2^{(M)}$and, …, and $x_9^{(M)} \in A_9^{(M)}$, then $y_M(x) = z_{M0} \sum_{j=1}^{9} (z_M x_j)$ .

To reduce the computational complexity, we assume that the number of rules

coincides with the number of membership functions, although they may differ.

In the first layer, fuzzification is performed according to formula (5) for each variable $x_i$. In this case, for each rule $R_j$, the values of the membership function $\mu_A{}^{(k)}(x_i)$are determined:

$$\mu_A{}^{(k)}(x_j) = \frac{1}{1+\left(\frac{\left(x_j - c_i{}^{(k)}\right)}{\sigma_i{}^{(k)}}\right)^{2b_i{}^{(k)}}}. \quad (7)$$

In the second layer, the coefficients ($\mathcal{W}_k = \mu_A{}^{(k)}(x)$ are determined by aggregating the values of the variables $x_i$. The ($\mathcal{W}_k$ coefficients are passed to the 3rd layer, where they are multiplied by the $y_i(x)$ values, and to the fourth layer to calculate the sum of the weights.

The third layer calculates the $y_i(x) = z_{k0} \sum_{j=1}^{N}(z_{kj} * x_j)$, values, which are multiplied by the $\mathcal{W}_k$ weight coefficients. The linear parameters $z_{k0}$ and $z_{kj}$ are functions of the consequences of the rules, and $z_{k0}$ is considered as the center of the membership function.

The fourth layer is represented by two neurons ($f_1$ and $f_2$) performing result aggregation:

$$f_1 = \sum_{k=1}^{M} \mathcal{W}_k y_k(x) = \sum_{k=1}^{M}\left[\left(\prod_{j=1}^{N} \mu_A{}^{(k)}(x_j)\right)c_k\right], \quad (8)$$

$$f_2 = \sum_{k=1}^{M} \mathcal{W}_k = \sum_{k=1}^{M}\left[\prod_{j=1}^{N} \mu_A{}^{(k)}(x_j)\right]. \quad (9)$$

The fifth normalizing layer is represented by one neuron, where the weights are normalized and the output function is calculated:

$$y(x) = \frac{f_1}{f_2} = \frac{\sum_{k=1}^{M} \mathcal{W}_k y_k(x)}{\sum_{k=1}^{M} \mathcal{W}_k} = \frac{\sum_{k=1}^{M}\left[\left(\prod_{j=1}^{N} \mu_A{}^{(k)}(x_j)\right)c_k\right]}{\sum_{k=1}^{M}\left[\prod_{j=1}^{N} \mu_A{}^{(k)}(x_j)\right]}. \quad (10)$$

The parameter values for the first and third layers are selected at the training stage. The parameters of the first layer $c_j{}^{(k)}, \sigma_j{}^{(k)}, b_j{}^{(k)}$ are considered nonlinear, and the parameters of the third layer $z_{kj}$ are considered linear. Training is performed in two steps. In the first step, the parameters of the membership functions of the third layer are selected by fixing individual parameter values and solving a system of linear equations

$$y(x) = \sum_{k=1}^{M} \mathcal{W}_k \, (z_{k0} + \sum_{j=1}^{N}(z_{kj} x_j)). \quad (11)$$

The output variables are replaced by reference values $d_p$ (P - is the number of training samples). The system of equations can be written in matrix form: $D_p = W * Z$. The solution to the system of equations is found by means of the pseudo-inverse matrix $W^+ : Z = W^+ D_p$. Next, after fixing the values of the linear parameters $z_{kj}$, the vector $Y$ of the actual output variables is calculated and the error vector $E = Y - D_p$ is determined.

At the second step, the errors are directed in the opposite direction to the first layer, where the parameters of the gradient vector of the target membership function are calculated relative to the parameters $c_j{}^{(k)}, \sigma_j{}^{(k)}, b_j{}^{(k)}$. Then, the parameters of the membership functions are adjusted using the fast gradient descent method:

$$c_j{}^{(k)}(n+1) = c_j{}^{(k)}(n) - \eta \frac{\partial E(n)}{\partial c_j{}^{(k)}}, \quad (12)$$

$$\sigma_j{}^{(k)}(n+1) = \sigma_j{}^{(k)}(n) - \eta \frac{\partial E(n)}{\partial \sigma_j{}^{(k)}}, \quad (13)$$

$$b_j{}^{(k)}(n+1) = b_j{}^{(k)}(n) - \eta \frac{\partial E(n)}{\partial b_j{}^{(k)}}, \quad (14)$$

where $n$ - is the iteration number; $\eta$ - is the learning rate parameter.

After specifying the nonlinear parameters, the process of adapting the linear and nonlinear parameters is started again. The iterative process is repeated until all process parameters are stabilized.

To estimate the probabilistic membership of a data packet to the clusters of
normal and atypical packets, the membership degree distribution is applied over three intervals:

a) 0–33 (%), b) 34–66 (%), c) 67–100 (%).

The decision to syn the size filtering rules is made when a packet falls into the third interval. Such a packet is considered atypical, the source is identified as a compromised node and its address is added to the exclusion rule of the flow table. If a packet falls into the first interval, it is considered normal. A rule is created for retransmitting packets with similar sets of characteristics and the source address is added to the white list. If a packet falls into the second interval, its degree of proximity to the clusters of normal and atypical packets is questionable. The point corresponding to the feature vector of such a packet becomes the centroid of a new cluster. The number of clusters increases by one, and the clustering problem is solved again. The result of intelligent analysis of data packets is the synthesis of rules for their filtering/retransmission in switches, inclusion of source addresses in white or black lists, adding records to the security event log, which include the identifier of the compromised node, IP address, attack detection time, event recording time, transaction type, etc.

## 3 Results

A subsystem of intelligent analysis and clustering has been developed, which includes modules for validating network nodes, filtering traffic, logging, a parser and packet analyzer based on fuzzy logic and a neural network, which is part of a system for secure Environmental Monitoring of critical events in the road transport infrastructure (Fig. 3).

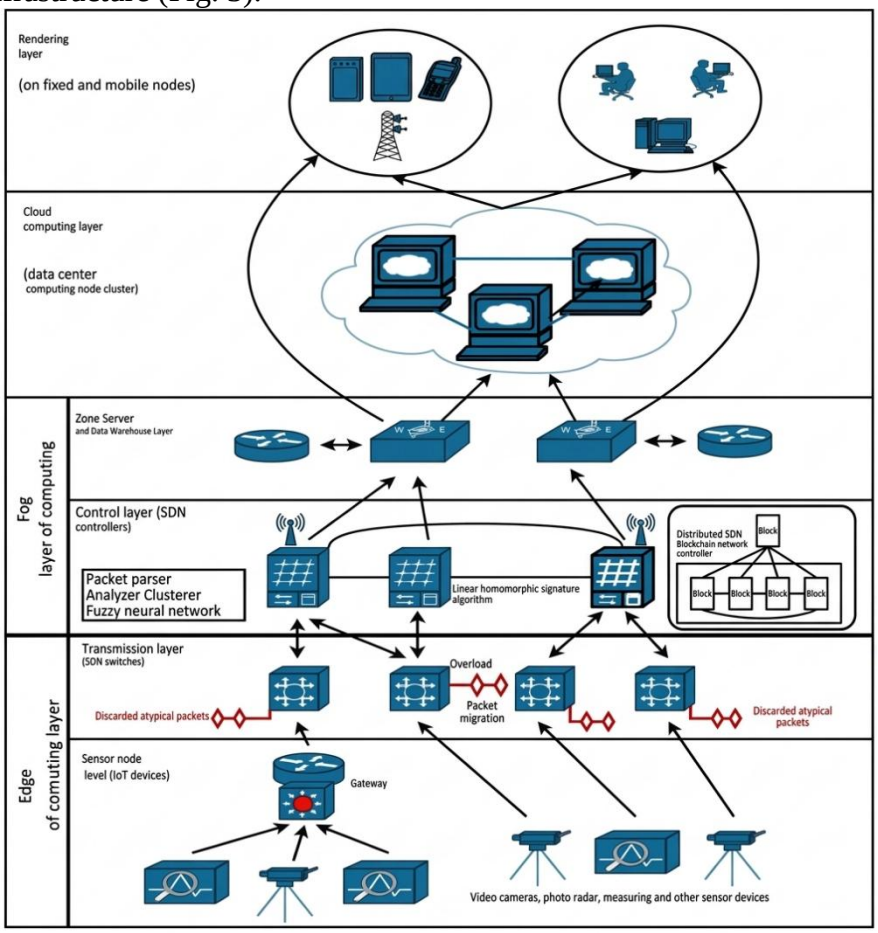



**Fig. 3.** Architecture of a secure Environmental Monitoring system

The system includes four layers with six data management levels:

1) edge computing layer, including the level of end sensor nodes, routers, sensor network coordinators and the level of data in SDN switches. The layer implements the processes of collecting, processing

and storing telemetry indicators of equipment operation and sensory data, and the function of packet relaying/filtering in SDN switches according to the rules of flow tables;

2) fog computing layer, including the level of zonal servers and the control level in SDN controllers. The layer implements the procedures for synthesizing flow table rules, validating the digital signature of network nodes, analyzing and filtering network traffic, detecting atypical packets based on header analysis and clustering, training a neural network, and logging. The zonal servers perform the functions of data consolidation, aggregation and storage based on the distributed hash table (DHT) technology, since the bulk of the data is photographs or videos from the scene of the incident, which solves the problem of storage scalability;

3) a cloud computing layer on the data center servers, which implements the main functionality of the Environmental Monitoring system for processing and predictive analysis of data, centralized storage of results;

4) a visualization layer is designed to present the Environmental Monitoring results on mobile and stationary user devices.

The modeling of the network architecture of the Environmental Monitoring system was performed in the Network Simulator version 3 (NS3) system [22, 23]. To evaluate the efficiency of packet analysis, several unauthorized nodes were created that generate atypical packets. The result of the modeling is estimates of the architecture performance indicators, such as transmission delay, response time, throughput and packet recognition accuracy. The growth of the transmission delay was estimated depending on the increase in the number of sensor nodes and the number of processed packets (Fig. 4).

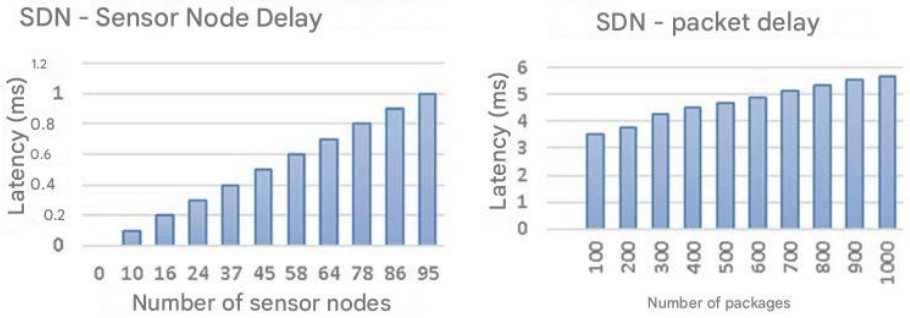



**Fig. 4.** Delay plots on the number of sensor nodes and the number of packets

The graphs show that the delay increases as the number of devices increases, with changes of several milliseconds after the packet is

transmitted from the end node. However, as the number of nodes increases, the delay becomes significant, as a linear dependence is observed.

The response time is determined based on requests received from end nodes to the gateway, which are transmitted to switches, where they are retransmitted or filtered after packet analysis. Comparison of the response time in the proposed architecture with the response time in the traditional architecture without analysis and clustering showed an increase in delay by an average of 10.5% (Fig. 5).

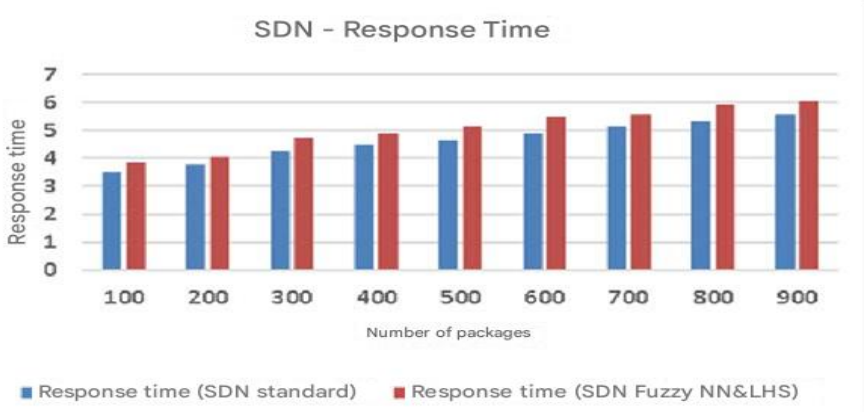



**Fig. 5.** Comparison of the response time of a network with a traditional SDN architecture and a network with an analysis and clustering subsystem

Fig. 6 shows diagrams of changes in throughput depending on the number of requests to network nodes. The reduction in throughput by an average of 16% in the proposed architecture occurs due to the time spent on analysis, recognition and clustering of packets.

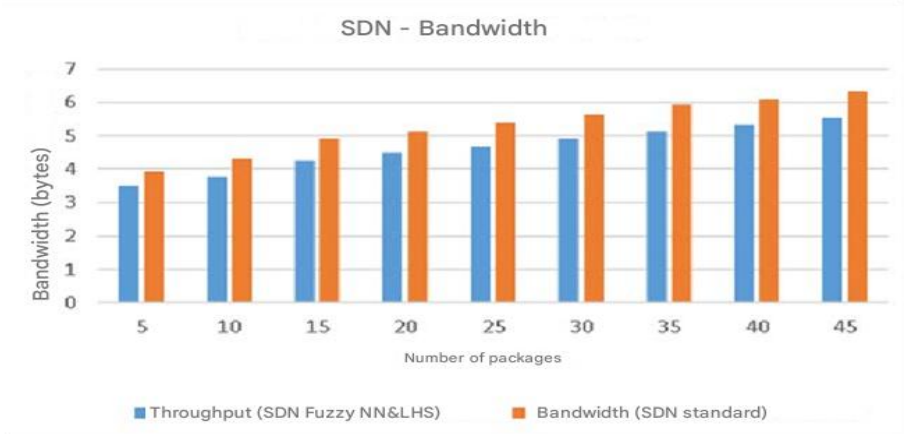



**Fig. 6.** Comparison of throughput in networks with traditional and proposed architecture

To evaluate the effectiveness of the proposed approach, an assessment of the accuracy of recognizing atypical packets is used. Table 1 shows the accuracy of detecting atypical packets depending on the rate of arrival of switches. The accuracy of detection is quite high, but decreases as the rate of arrival of packets increases, which is due to the time of packet analysis and synthesis of flow table rules.

**Table 1.** Accuracy of detecting atypical packets.

| **Packets received for processing (packets/sec)** | **Total number of packages** | | **Number of packets as a result of clustering** | | **Recognition accuracy (%)** |
|---|---|---|---|---|---|
| | Normal packages | Atypical packages | Normal packages | Atypical packages | |
| **20** | 400 | 75 | 399 | 76 | **98,68** |
| **40** | 800 | 150 | 797 | 153 | **98,03** |
| **60** | 1200 | 225 | 1195 | 230 | **97,82** |
| **80** | 1600 | 300 | 1590 | 310 | **96,77** |
| **100** | **2000** | **375** | **1980** | **395** | **94,94** |

## 4 Conclusion

During the research, a secure architecture of the road traffic Environmental Monitoring system was synthesized using SDN network technology, distributed registry, analysis and clustering of data packets using a fuzzy neural network. The paper presents a number of results:

a) SDN architecture of the system with four device layers and six data management levels;

b) a method for detecting cyberattacks using a fuzzy neural network and a clustering algorithm for evaluating packets by nine characteristics;

c) a set of data packet characteristics for recognizing and probabilistically evaluating malicious packets.

Modeling and evaluation of the effectiveness of the proposed approach showed a fairly high speed of traffic processing in the Environmental Monitoring system.

To reduce the computational load, a three-stage analysis technology is implemented. At the first stage, invalid traffic is filtered by port numbers. At the second stage, the authenticity of data sources and access devices is confirmed and traffic from invalid nodes is filtered. At the third stage, machine learning procedures and intelligent analysis of filtered packets are performed.

## References


1. de Campos Souza, P.V., Guimarães, A.J., Rezende, T.S., Silva Araujo, V.J., Araujo, V.S.: Detection of Anomalies in Large-Scale Cyberattacks Using Fuzzy Neural Networks. AI 1(1), 92–116 (2020).
2. Moshkin, V., Kurilo, D., Yarushkina, N.: Integration of Fuzzy Ontologies and Neural Networks in the Detection of Time Series Anomalies. Mathematics 11(5), 1204 (2023).
3. Ahmad, Z., Shahid Khan, A., Nisar, K., Haider, I., Hassan, R., Haque, M.R., Tarmizi, S., Rodrigues, J.J.P.C.: Anomaly Detection Using Deep Neural Network for IoT Architecture. Applied Sciences 11(15), 7050 (2021).
4. Souza, P.V., et al.: Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling. Journal of Parallel and Distributed Computing 73(6), 790–806 (2013).
5. Lee, W., Maidanov, K.: Fog-Assisted Anomaly Detection in IoT-WSN Ecosystems Using Hybrid Deep Learning Models. Journal of Wireless Sensor Networks and IoT 3(1), 1–9 (2025).
6. Ahmad, Z., et al.: Deep neural network-based anomaly detection for IoT network flows. Applied Sciences 11, 7050 (2021).
7. Hussein, S.F., Dallalbashi, Z.E., Mohammed, A.B.: Anomaly Detection in Internet of Medical Things with Artificial Intelligence. Eastern-European Journal of Enterprise Technologies (2023).
8. Atassi, R.: Anomaly Detection in IoT Networks: Machine Learning Approaches for Intrusion Detection. Journal of Intelligent Systems and Internet of Things 13(1), 126–134 (2023).
9. Mardenov, Y., Adamova, A., Zhukabayeva, T., Othman, M.: Enhancing Fault Detection in Wireless Sensor Networks Through Support Vector Machines. Journal of Robotics and Control 4(6) (2024).
10. Xiaofeng, W.: Federated deep learning for anomaly detection in the Internet of Things. Computers & Electrical Engineering (2023).
11. Karthikeyan, M., Manimegalai, D., RajaGopal, K.: Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection. Scientific Reports 14, Article 231 (2024).
12. Bashar, A., Hanif, M., Alshammari, N., Alqahtani, A., Alshahrani, M.: A novel hybrid technique using fuzzy logic, neural networks and genetic algorithm for intrusion detection system. Measurement: Sensors 30, 100933 (2023).
13. Abirami, S., Chitra, P.: A novel enhanced neural network for anomaly detection in the IoT environment. Computers & Electrical Engineering 129(B), 110833 (2026).
14. Hamed, B., Ali, A., Mohamed, E.: A fuzzy neural network approach for online fault detection in waste water treatment process. Computers & Electrical Engineering 40(7), 2216–2226 (2014).
15. Caville, E., Lo, W.W., Layeghy, S., Portmann, M.: Anomal-E: A Self-Supervised Network Intrusion Detection System based on Graph Neural Networks. arXiv (2022).
16. Nguyen, T.D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., Sadeghi, A.R.: DÏoT: A Federated Self-learning Anomaly Detection System for IoT. arXiv (2018).
17. Lo, W.W., Layeghy, S., Sarhan, M., Gallagher, M., Portmann, M.: E-GraphSAGE: A Graph Neural Network based Intrusion Detection System for IoT. arXiv (2021).
18. Sáez-de-Cámara, X., Flores, J.L., Arellano, C., Urbieta, A., Zurutuza, U.: Clustered Federated Learning Architecture for Network Anomaly Detection in Large Scale Heterogeneous IoT Networks. arXiv (2023).

19. Kasabov, N.: Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering. MIT Press, Cambridge (1996).
20. Kasabov, N.: Evolving Connectionist Systems: The Knowledge Engineering Approach. Springer, London (2007).
21. Kasabov, N.: Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence. Springer, Berlin (2019).
22. Halder, A., Konar, A., Mandal, R., Chakraborty, A., Bhowmik, P.: General and Interval Type-2 Fuzzy Face-Space Approach to Emotion Recognition. IEEE Transactions on Systems, Man, and Cybernetics (2013).
23. Saha, A., Konar, A., Nagar, A.K.: EEG Analysis for Cognitive Failure Detection in Driving Using Type-2 Fuzzy Classifiers. IEEE Transactions on Emerging Topics in Computational Intelligence (2017).

# The Problem of Forming a Feature Space for Speaker Recognition

Nomaz Mirzaev[1][0000-0002-0657-2654], Johongir Urinboev[2][0009-0000-2305-8887],

Sayyora Ibragimova[2][0000-0002-8118-8919], Gulmira Mrzaeva[1][0000-0002-0412-3951],

Azizbek Tillavoldiev[2][0009-0004-7579-2185]

[1] Tashkent university of information technologies named after Muhammad al-Khwarizmi Tashkent, Uzbekistan
[2] Digital Technologies and Artificial Intelligence Research Institute Tashkent, Uzbekistan
jahongir8010@gmail.com

**Abstract.** This work examines the issue of person identification based on voice signals. Speech-based biometric systems are widely used in information security, human-computer interaction, and intelligent information systems. However, noise, channel effects, and high dimensionality of features can negatively affect the system's accuracy. This study investigated the development of noise-resistant and computationally efficient recognition algorithms based on a voice dataset consisting of 50 English speakers. Acoustic features such as MFCC, LPC, PLP, spectral centroid, energy, and entropy were extracted from speech signals, and a single feature vector was formed through feature fusion of these features. In this research, person recognition methods were thoroughly studied, with the main focus on two important approaches: feature fusion and feature space reduction using Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Extensive experiments were conducted on various speech datasets characterized by different noise levels and numbers of speakers. The study yielded good results for a single dataset and classifiers. To reduce redundant and mutually correlated parts of features, dimensionality reduction methods - PCA and ICA - were applied to the feature space. The resulting features were evaluated using K-Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA) classifiers. These results indicate a significant increase in computational speed and speaker recognition efficiency due to the feature fusion and feature space reduction approaches.

**Keywords** Speaker identification, speaker verification, feature extraction, feature fusion, feature space reduction, PCA, ICA, KNN, LDA.

## 1 Introduction

This work presents an extended analysis of speaker recognition (SR) systems, focusing on feature fusion methods applied to speech datasets of various sizes [1]. In the approach proposed in this study, the model integrates 18 distinct features via feature

fusion, including mel-frequency cepstral coefficients (MFCC), linear predictive coding (LPC), perceptual linear prediction (PLP), root mean square (RMS) value, centroid and entropy features, as well as their corresponding delta (Δ) and delta-delta (ΔΔ) feature vectors. Experimental results demonstrated that feature fusion of different features together with their associated delta and delta-delta components leads to a significant increase in speaker recognition accuracy. This resulted in improved recognition rates for noise-free voice data using linear discriminant (LD) and K-nearest neighbor (KNN) classifiers. Furthermore, the Equal Error Rate (EER) for speaker verification was reduced compared to using a single feature.

Feature fusion of numerous features may appear beneficial for capturing diverse aspects of the data; however, it can introduce dimensionality-related challenges, including increased computational complexity, overfitting, and reduced efficiency. In the effective design of speaker recognition systems, it is crucial to balance the number of features with the desired performance indicators and computational constraints. To address this issue, methods for reducing the feature space are employed, including Principal Component Analysis (PCA) and Independent Component Analysis (ICA). These techniques compress data by reducing dimensionality while preserving the most important features. Consequently, the training process accelerates and computational efficiency improves [2,3].

The main achievements of this work are as follows:

• We have conducted a detailed study of methods for reducing the feature space, specifically designed to address the complexities of high-dimensional data in speaker recognition systems.

• Overall, the research emphasizes the importance of feature fusion in speaker recognition systems and demonstrates the advantages of feature space reduction methods.

• The aim is to achieve an optimal feature fusion strategy that not only improves recognition accuracy but also reduces the dimensionality of the feature space and enables faster computation.

• Our proposed models provide reliable solutions for improving the efficiency of recognizing various numbers of speakers in noisy environments, using English-language speech datasets containing 50 speakers. Our models are adaptable and can be applied to a wide range of application domains and datasets with different scales and characteristics.

## 2 Problem Statement

The primary objective of the research is to enhance the accuracy and computational efficiency of speaker recognition algorithms by integrating various acoustic features extracted from voice signals and reducing their dimensionality using PCA and ICA methods.

- To achieve this goal, the following tasks were established:
- extracting MFCC, LPC, PLP, and additional spectral features from speech signals;
- fusing the features into a single feature-level vector;
- reducing the dimensionality of features using PCA and ICA methods;

- evaluating identification accuracy using KNN and LDA classifiers;
- comparing and analyzing the obtained results.

# 3 Proposed methodology

## 3.1 Exploration of the problem

Our primary goal is to develop the most optimal set of features that enhance accuracy, reduce errors, and improve computational efficiency, thereby conserving computational resources. Moreover, we aim to create a universal approach that can be applied to datasets of varying sizes. In the field of speaker recognition, relatively few studies have focused on computational time. Our proposed work addresses this gap by providing detailed information about computational time, which is a crucial factor alongside accuracy and error rates. To achieve this objective, we present two distinct approaches for speaker identification (SI).

1. Methodology for feature fusion: Feature fusion, which transforms features from various sources or databases into a single, enriched feature set, is considered a fundamental approach in speaker recognition systems. Spectral features, encompassing frequency, power (energy), and other signal properties such as MFCC, LPC, PLP, spectral centroid, and entropy, provide a stable foundation. On the other hand, prosodic features include auditory characteristics such as stress, pitch variation, and intonation (e.g., RMS). Although spectral features, especially MFCC, have demonstrated higher performance compared to pitch-based systems (e.g., pitch, RMS), their combined use provides unparalleled stability, which is crucial for recognition systems. Furthermore, feature derivatives enable the quantitative assessment of subtle changes in voice signals. We aimed to increase accuracy by extracting six features and calculating their derivatives and second-order derivative values [6-9].

2. Feature space reduction: This approach focuses on reducing the dimensionality of the feature set, optimizing computational processes, while preserving crucial information for accurate speaker recognition. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are applied in our work to reduce the feature space. However, it should be noted that feature fusion is a complex process that may slow down the computational process. Therefore, carefully considering the balance between PCA and ICA is important to ensure equilibrium between computational efficiency and information retention [13-16].

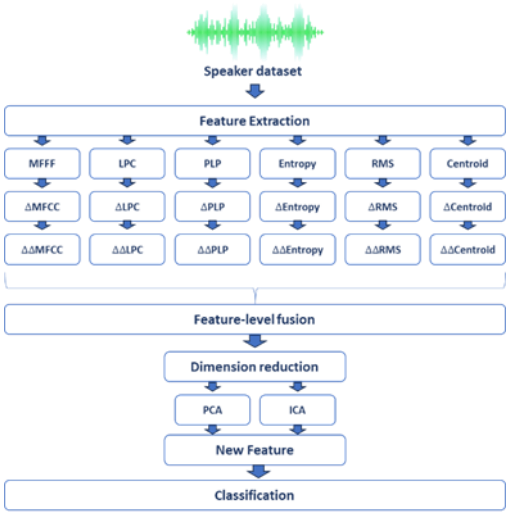



**Fig. 1.** Proposed approach for the speaker recognition model.

### 3.2 The approach to feature fusion (Approach 1)

Feature fusion is the process of harmonizing features extracted from various sources or databases to create a comprehensive and enriched set of features. This integration enhances the distinctiveness of speech features, thereby facilitating more accurate classification and speaker recognition [1]. Figure 2 illustrates the approach to feature fusion applied to voice data.

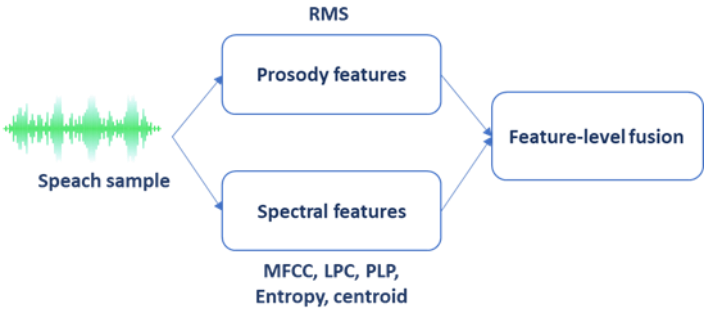



**Fig. 2.** Method of feature fusion

Methods of feature extraction. Mel-frequency cepstral coefficients (MFCC) [17; 18] are often used in automatic speech recognition [19] and are generated by correlating filter banks based on a Type 2 discrete cosine transform. They were introduced in the 1980s by Davis and Mermelstein, with the idea stemming from an attempt to preserve important phonetic information from speech that may have syntactic changes and variations in duration. MFCC extracts features that are more independent of such variations [20]. Mel-scale filter banks have a high degree of correlation, as triangular filters partially overlap each other. Applying the discrete cosine transform to each filter bank yields the Mel-frequency cepstral coefficients (MFCC). Usually, half of them are discarded, as higher-numbered filter banks indicate rapid changes in high

energy. Mel-frequency cepstral coefficients are less susceptible to additive interference. Based on psychoacoustic theory [11], MFCC features are widely used in speaker recognition systems. The $i$-vector speaker recognition system proposed by Dehak et al. [12] considers MFCCs as the primary source of speech features.

The selection of the primary 13 DCT coefficients as mel-frequency cepstral coefficient (MFCC) features by converting mel-scale frequencies to a logarithmic scale and applying discrete cosine transform (DCT) is established in these works [10,11,21-23]. Linear Predictive Coding (LPC) is a method of computing the current sample by linearly combining previous samples, where inverse filtering is applied to eliminate formants from speech signals, resulting in a residual signal. To calculate LPC coefficients, the VQ-LBG algorithm is used, in which vector quantization (VQ) of LPC coefficients is performed in the Linear Spectral Frequency (LSF) domain to reduce the data transmission rate.

Understanding the autoregressive (AR) version of speech is crucial for determining LPC. The sound signal is modeled as a $p$-order AR technique, which is expressed by equation (1), where each sample at the $n$-th moment is influenced by $p$ previous samples and $u(n)$ Gaussian noise. The LPC coefficients are denoted by $a$ and are determined using the Yule-Walker equations described in equations (2) - (4), with the solution given in expression (5). For simplicity, only the first 13 LPC coefficients are used. More detailed information about the proposed LPC features, VQ-LBG algorithm, and calculation steps can be found in works [24,25]. Linear predictive coding (LPC) is indeed a powerful method, but it may have limitations, particularly sensitivity to changes in sampling rate and reliance on assumptions about speech signal characteristics. Therefore, in our proposed work, we complement LPC with a number of additional features to enhance the robustness and effectiveness of the speaker recognition system.

$$s(n) = -\sum_{k=1}^{p} a_k \times (n-k) + u(n) \tag{1}$$

$$D(i) = a_0 + \sum_{n=1}^{N} \left( s(n) \times (n-i) \right) \tag{2}$$

The final form of the Yul-Walker expressions is represented by equations (3) and (4).

$$\sum_{k=1}^{p} a_k D\left(|i-k|\right) = -D(i) \tag{3}$$

$$\begin{bmatrix} \mathrm{D}(0) & \cdots & \mathrm{D}(\mathrm{p}-1) \\ \vdots & \ddots & \vdots \\ \mathrm{D}(\mathrm{p}-1) & \cdots & \mathrm{D}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} = - \begin{bmatrix} \mathrm{D}(\mathrm{i}) \\ \vdots \\ \mathrm{D}(\mathrm{K}) \end{bmatrix} \tag{4}$$

Expression (5) provides the final solution for determining the LPC coefficients.

$$a = -D^{-1} r \tag{5}$$

In this research, PLP markers are selected due to their effectiveness in noise reduction, reverberation suppression, and echo cancellation, which increases overall performance efficiency. Extracting PLP features involves several steps, including pre-

emphasis, cube root compression, and removal of unnecessary speech data [5,26,27]. Thirteen PLP features are determined by calculating the average of all PLP features for each voice, which effectively reduces system complexity using PYTHON software [26,27]. To match the dimensions of PLP features with the dimensions of other features, the average value of each frame is calculated, resulting in 13×1 feature vectors in each frame. Figure 2 provides a visual representation of the feature extraction process for MFCC, LPC, and PLP features.

The spectral centroid (SC) determines the center of gravity of the spectrum, which is considered important, and represents a unique value that expresses the frequency domain characteristic of the speech signal. A high SC value indicates that the signal is strong [28], which is calculated using expression (6). The variable $x_i$ represents the $i$ -th part of the speech signal, while $x_i k$ denotes the amplitude value in the $k$-th frequency interval of the $i$-th part of the speech signal.

$$C(i) = C_1(i) / C_2(i) \tag{6}$$

$$C_1(i) = \sum_{k=0}^{N-1} k\left|x_i(k)\right|, C_2(i) = \sum_{k=0}^{N-1}\left|x_i(k)\right| \tag{7}$$

Spectral entropy (SE). To calculate entropy as a feature, the following steps are carried out:

• We calculate the power spectral density $s(f)$ for the signal $x(t)$.

• Based on the selected frequency range, we calculate the power of the spectral range. After calculating the power of the spectral range, we normalize the power in the selected range.

• We calculate the spectral entropy using expression (8) [29].

$$SE = \sum s(f) \times \ln \frac{1}{s(f)} \tag{8}$$

The root mean square (RMS) is a measure that indicates the amplitude of an audio signal. It is calculated by taking the square root of the sum of the mean squares of the sound sample amplitudes. The RMS formula is given in expression (9) [30]. Here, $x_1, x_2, \ldots x_n$ represent $n$ observations, and $x_{rms}$ denotes the RMS value for these $n$ observations.

$$x_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{9}$$

Delta features are crucial for reflecting the rate of change in voice features, especially the power dynamics of speech signals relative to noise. Valuable dynamic information is extracted from speech signals using delta (Δ) and delta-delta (ΔΔ) features [4,9,23]. The calculation of the $\Delta z$ feature is carried out by subtracting the previous feature $f_{z-1}$ from the current feature $f_z$, as shown in equation (10).

$$\Delta z = f_z - f_{z-1} \tag{10}$$

Similarly, the feature $\Delta\Delta z$ is obtained by subtracting the current feature $\Delta z$ from the previous feature $\Delta\Delta_{z-1}$, as shown in equation (11).

$$\Delta\Delta z = \Delta z - \Delta_{z-1} \tag{11}$$

Feature fusion method. The following steps were taken to select a model for feature fusion:
1. All 18 features were tested separately using speech data from 50 speaking English.
2. Out of the 18 features, the 2 best features with the highest SI accuracy and the lowest average EER value were selected.
3. When determining the best model, the average accuracy and average EER values across three classifiers were taken into account. LPC and PLP were identified as the first and second best features, respectively. Equation (12) shows the method for calculating the average accuracy based on the results of all three classifiers.

$$Average\,accuracy\,resault = \frac{\mathrm{LD} + \mathrm{KNN}}{2} \tag{12}$$

4. In the second stage, the best LPC and PLP features are separately fused with the remaining 17 features. Two more best models are selected from this process. The two best models identified at this stage were the combination model of MFCC and LPC, and the combination model of PLP and LPC.
5. In the third stage, the remaining 16 features are separately combined with the two best models selected from the second stage, resulting in a three-feature fusion configuration.
6. Feature fusion configurations involving 4 to 18 features are performed in the same manner, and at each stage, the two best-performing models are selected. A total of 315 models are tested for the dataset.

### 3.3 Methods of reducing the feature space (Approach 2)

To address the computational challenges associated with feature fusion in Approach 1, specifically to simplify the time-consuming and labor-intensive process of identifying the best models, a more efficient solution is needed. To overcome these issues, feature space reduction algorithms have been developed, aimed at simplifying calculations by reducing the size of training samples.

**Principal Component Analysis (PCA).** One of the widely used methods in this regard is Principal Component Analysis (PCA). PCA addresses the complexity problem arising in high-dimensional space by reducing the dimensions while preserving the components that represent the maximum variance of the data. In high-dimensional spaces, data points are usually sparse, making them difficult to generalize and study. PCA identifies the directions (principal components) in which the data vary the most and projects the data onto a lower-dimensional subspace covered by these components. This reduces the dimensionality of the space while retaining most of the important information in the data.

The steps performed in PCA are as follows [31]:

1. Loading input data: The feature fusion model, which serves as the input dataset, is fed into the PCA algorithm.

2. Subtracting the mean: The mean value of the data is subtracted from each feature in the initial dataset. This step ensures the centralization of data around the origin of the coordinate system.

3. Calculating the covariance matrix: The covariance matrix of the dataset is calculated. This matrix reflects the relationships and variations between different features.

4. Determining eigenvectors: The eigenvectors associated with the largest eigenvalues of the covariance matrix are identified. These eigenvectors represent the directions of maximum variance in the dataset.

5. Projecting the dataset: The initial dataset is projected onto the eigenvectors obtained in the previous step. This projection transforms the data into a lower-dimensional subspace spanned by the eigenvectors.

6. By following these steps, PCA effectively reduces the dimensionality of the dataset while preserving the most important information and capturing the most significant variations in the data [31,32].

**Independent Component Analysis (ICA).** ICA was first proposed in the 1980s by J. Herault, C. Jutten, and B. Ans, who developed an iterative algorithm for real-time processing [33]. Independent Component Analysis (ICA) is a method of dimensionality reduction in feature space, aimed at extracting independent components from a dataset. It is an extended version of PCA, allowing the identification of latent factors or sources that contribute to the formation of observed data. ICA is widely used in signal processing and data analysis due to its ability to separate mixed signals into their original sources.

The main steps performed in ICA can be summarized as follows:

1. Preprocessing: As in PCA, data is usually preprocessed - centered to provide a common reference point, and features are scaled to ensure equal contribution from each feature.

2. Whitening: Whitening is performed to transform the data into a new representation, where features are uncorrelated with each other and their variance is set to 1. This step helps eliminate any linear dependencies between features.

3. Determination of non-Gaussianity criterion: The ICA method aims to find components that are statistically as independent as possible. Various non-Gaussianity measures, such as kurtosis or negentropy, are used to quantify the degree of deviation from Gaussianity and to guide the process of separating independent components.

4. Optimization: The main task of ICA is to maximize the non-Gaussianity criterion for each component. This is achieved through an optimization process that involves finding weights or mixing matrices that maximize the non-Gaussianity criterion.

5. Iterative estimation: ICA often involves an iterative estimation process to better separate independent components.

6. Reconstruction: After independent components are obtained, they can be used to reconstruct the original dataset or further analyzed for specific purposes, such as feature extraction or signal separation [32-34].

These approaches to reducing the feature space help identify the best models in the integration process more quickly and effectively.

### 3.4 Classification

**Linear Discriminant Analysis (LDA) Classifier.** The Linear Discriminant Analysis (LDA) classifier is a statistical classification method based on finding the linear projection that best reveals the differences between classes. Its principle is to project data from a given feature space in such directions that the within-class scatter is minimized while maximizing the between-class separation. In practice, LDA relies on the mean vector and pooled covariance for each class, and the decision boundary takes the form of a hyperplane (i.e., the classification is "linear"). Consequently, LDA is computationally efficient, performs robustly on small and medium-sized datasets, and yields particularly good results when classes are approximately normally distributed with similar covariances. However, LDA's capabilities may be limited if the data are strongly nonlinearly separable or if class covariances differ significantly. Comprehensive information about LDA is provided in reference [35].

**K-Nearest Neighbors (KNN) Classifier.** The K-Nearest Neighbors (KNN) approach is a method of classifying unknown data points based on their similarity to neighboring data points. In this approach, the parameter K indicates the number of dataset elements that contribute to the classification process, and it is set to K=5 for this specific work. The KNN algorithm can be summarized in the following steps [36]:

1. We select the value of K, which represents the number of neighbors.
2. We calculate the Euclidean distance between the unknown given point and its K nearest neighbors.
3. We classify the K nearest neighbors according to the calculated Euclidean distances.
4. We count the number of data points in each class.
5. We assign the new data point to the class with the highest count.

## 4 Evaluation

### 4.1 Database preparation

In the proposed work, we used a dataset consisting of 50 English speakers. The dataset comprises 2,511 speech samples from 50 speakers in total. For classification studies, the dataset was divided into training (70%) and testing (30%) subsets using stratified sampling. We used 1,758 samples for the training set and 753 samples for the testing set. The average duration of each sample was approximately 59 seconds, with the total duration of speech being approximately 41 hours. All audio files are stored in WAV format and recorded in mono. The recordings have sampling frequencies of 16 kHz and 22.05 kHz; however, to ensure accuracy in feature extraction, all signals were resampled to 16 kHz.

### 4.2 Evaluating the Effectiveness of Speaker Identification (SI)

Speaker recognition accuracy evaluates the system’s ability to correctly identify different speakers based on their voices. This is crucial as it provides security, personali-

zation, efficiency, and convenience for applications that utilize voice communication or authentication. The accuracy of Speaker Identification (SI), calculated using PYTHON's classification learning library, encompasses all proposed models with both classifiers. Simplified results tables only reflect the models with the highest accuracy. Accuracy is determined by the ratio of correctly identified speakers voices to the total number of voices in the dataset (equation 13).

$$Accuracy = \frac{\text{Number of voices correctly identified}}{\text{Total number of audio files}} \tag{13}$$

### 4.3 Evaluating the Effectiveness of Speaker verification (SV)

For the SV problem, EER values are calculated using PYTHON for each model. The EER is determined by the intersection point of the receiver operating characteristic (ROC) curve and the diagonal axis from (0, 1) to (1, 0), which corresponds to the false positive rate [37, 38].

Figure 3 shows the ROC curve and EER point for individual MFCC and LPC features, as well as feature fusion configurations—including fusion of all features and Δ-fusion—constructed by isolating and pairing their feature vectors in various combinations. The figure also reports the results of feature-space reduction using PCA and ICA, with all experiments evaluated using the KNN classifier. We observe that the ROC curve closer to the point (1, 0) yields the lowest EER, indicating better performance compared to other applied methods. Thus, based on this curve, we can conclude that among all the methods used, the feature space reduction method using the ICA technique produces the best results.

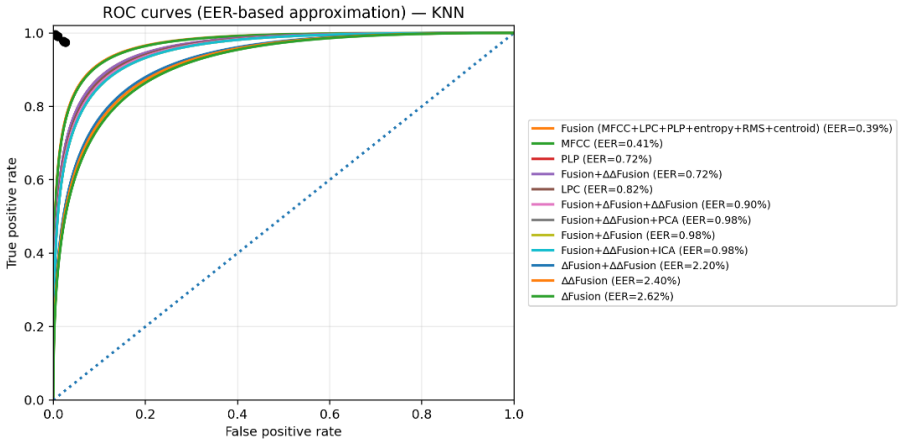



**Fig. 3.** ROC curve for English speech data from 50 speakers using the KNN classifier

Figure 4 depicts the ROC curve and EER point for various feature vectors using the LDA classifier. These include individual MFCC, PLP, and LPC features, as well as feature fusion configurations—such as fusion of all features, Δ-fusion, and their various combinations—together with feature-space reduction methods based on PCA and ICA. We observe that the ROC curve closer to the point (1, 0) yields the lowest EER, indicating better performance compared to other applied methods. Thus, based on this

curve, we can conclude that among all the methods used, the feature space reduction method using ICA technique produces the best results.

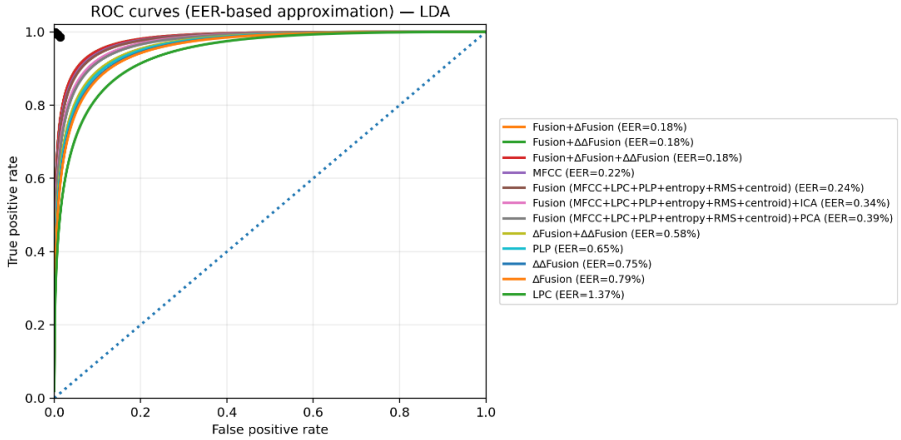



Fig. 4. ROC curve for English speech data from 50 speakers using the LDA classifier

## 5 Results Discussion

### 5.1 Optimal results were achieved through feature fusion (Approach 1)

Tables 1-2 reflect the results of speaker recognition using individual features such as MFCC and LPC, as well as their associations and performance indicators of the most effective feature fusion model. This model achieves the highest accuracy and the lowest equivalent error level (EER) in experiments conducted on a dataset of 50 speakers. The presented data demonstrates the advantages of feature fusion, which is consistently preferable to relying on a single feature or a limited number of fusion configurations. Tables 1-2 include only the classification method that yielded the best results.

In the experiment, the results achieved through the feature fusion approach showed that the best Speaker Identification (SI) accuracy and Equal Error Rate (EER) values obtained using the Linear Discriminant (LD) classifier were 97.4% accuracy with an EER value of 0.18% for 50 speakers. Similarly, the best SI accuracy and EER values obtained using the K-Nearest Neighbors (KNN) classifier were 95.82% accuracy with an EER value of 0.39% for 50 speakers. Below, we present the classification results using the feature fusion approach for the dataset of 50 speakers (Tables 1-2).

**Table 1.** Comparison of results using the feature fusion approach, with classification implemented in the LDA classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| MFCC | LDA | 0.026 | 0.0004 | 97.01 | 0.22 |
| LPC | LDA | 0.023 | 0.0003 | 94.22 | 1.37 |

| | | | | | |
|---|---|---|---|---|---|
| PLP | LDA | 0.065 | 0.006 | 95.8 | 0.65 |
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid) | LDA | 0.028 | 0.0004 | 97.01 | 0.24 |
| ΔFusion | LDA | 0.029 | 0.0003 | 93.03 | 0.79 |
| ΔΔFusion | LDA | 0.029 | 0.0004 | 93.43 | 0.75 |
| Fusion+ΔFusion | LDA | 0.047 | 0.0005 | 97.21 | 0.18 |
| Fusion+ΔΔFusion | LDA | 0.046 | 0.0005 | 97.21 | 0.18 |
| ΔFusion+ΔΔFusion | LDA | 0.045 | 0.0005 | 93.83 | 0.58 |
| Fusion+ΔFusion+ΔΔFusion | LDA | 0.082 | 0.0005 | 97.4 | 0.18 |

**Table 2.** Comparison of results using the feature fusion approach, with classification implemented in the KNN classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| MFCC | KNN | 0.002 | 1.355 | 95.82 | 0.41 |
| LPC | KNN | 0.002 | 0.015 | 95.42 | 0.82 |
| PLP | KNN | 0.004 | 0.08 | 94.9 | 0.72 |
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid) | KNN | 0.002 | 0.013 | 96.81 | 0.39 |
| ΔFusion | KNN | 0.0021 | 0.016 | 85.28 | 2.62 |
| ΔΔFusion | KNN | 0.0021 | 0.017 | 84.48 | 2.40 |
| Fusion+ΔFusion | KNN | 0.0034 | 0.048 | 94.62 | 0.98 |
| Fusion+ΔΔFusion | KNN | 0.0034 | 0.017 | 94.42 | 0.72 |
| ΔFusion+ΔΔFusion | KNN | 0.0032 | 0.015 | 87.27 | 2.20 |
| Fusion+ΔFusion+ΔΔFusion | KNN | 0.0037 | 0.019 | 93.43 | 0.90 |

Figure 5 illustrates how speaker recognition accuracy varies with different feature fusion sizes (i.e., different numbers of fused features) across multiple classification methods. Analysis of the graphs shows that increasing the number of features generally improves accuracy. However, it should be emphasized that randomly adding features does not guarantee better results. Furthermore, different feature fusion configurations exhibit varying performance across classifiers. These methods aim to identify the most informative features and reduce redundancy, maximizing the discriminative ability of the feature set. By reducing the feature space, we can improve the feature fusion process, thereby enhancing the robustness and effectiveness of speaker recognition systems under various noise conditions and across datasets of different sizes.

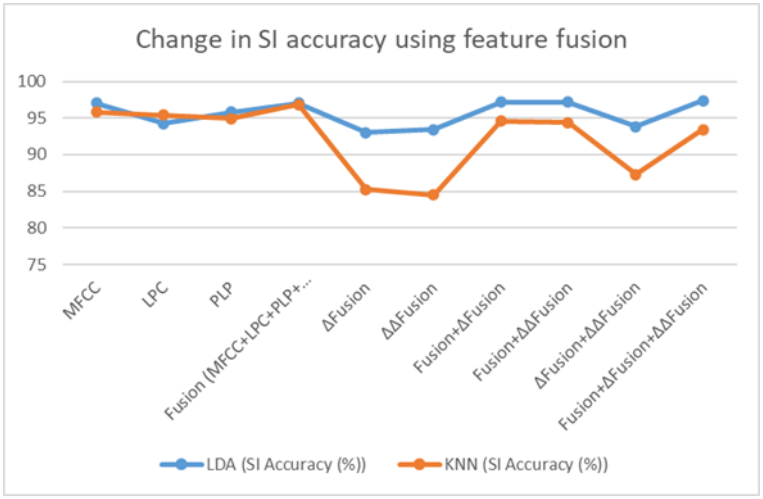



Fig. 6. Graph showing the change in SI accuracy when using feature fusion.

### 5.2 Optimal results using the feature space reduction approach (Approach 2)

Tables 3-6 present the best results achieved when applying the PCA and ICA methods for feature space reduction, including the computation times for both the training and testing phases.

**Table 3.** Comparison of results using the PCA method for feature space reduction. Classification was performed using the LDA classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid)+PCA | LDA | 0.02 | 0.0003 | 98.60 | 0.39 |
| ΔFusion+PCA | LDA | 0.02 | 0.0003 | 96.22 | 0.41 |
| ΔΔFusion+PCA | LDA | 0.03 | 0.0003 | 96.02 | 0.44 |
| Fusion+ΔFusion+PCA | LDA | 0.03 | 0.0003 | 97.81 | 0.37 |
| Fusion+ΔΔFusion+PCA | LDA | 0.03 | 0.0003 | 97.81 | 0.46 |
| ΔFusion+ΔΔFusion+PCA | LDA | 0.03 | 0.0003 | 96.42 | 0.37 |
| Fusion+ΔFusion+ΔΔFusion+PCA | LDA | 0.02 | 0.0003 | 97.61 | 0.22 |

**Table 4.** Comparison table of results using the PCA method for feature space reduction, with classification performed using the KNN classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid)+PCA | KNN | 0.001 | 0.01 | 96.62 | 1.72 |

| | | | | | |
|---|---|---|---|---|---|
| ΔFusion+PCA | KNN | 0.001 | 0.014 | 94.23 | 1.18 |
| ΔΔFusion+PCA | KNN | 0.001 | 0.014 | 93.43 | 1.11 |
| Fusion+ΔFusion+PCA | KNN | 0.001 | 0.012 | 95.82 | 1.28 |
| Fusion+ΔΔFusion+PCA | KNN | 0.001 | 0.013 | 97.01 | 0.98 |
| ΔFusion+ΔΔFusion+PCA | KNN | 0.001 | 0.013 | 95.62 | 0.97 |
| Fusion+ΔFusion+ΔΔFusion+PCA | KNN | 0.001 | 0.012 | 96.42 | 0.79 |

**Table 5.** Comparison of results using the ICA method for feature space reduction. Classification was performed using the LDA classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid)+ICA | LDA | 0.03 | 0.0003 | 98.40 | 0.34 |
| ΔFusion+ICA | LDA | 0.03 | 0.0003 | 96.42 | 0.40 |
| ΔΔFusion+ICA | LDA | 0.03 | 0.0003 | 96.02 | 0.41 |
| Fusion+ΔFusion+ICA | LDA | 0.03 | 0.0003 | 97.21 | 0.22 |
| Fusion+ΔΔFusion+ICA | LDA | 0.03 | 0.0003 | 97.61 | 0.47 |
| ΔFusion+ΔΔFusion+ICA | LDA | 0.03 | 0.0003 | 96.81 | 0.55 |
| Fusion+ΔFusion+ΔΔFusion+ICA | LDA | 0.03 | 0.0003 | 97.61 | 0.22 |

**Table 6.** Comparison of results using the ICA method for feature space reduction. Classification was performed using the LDA classifier.

| Feature used (Model) | Classifier Model | Training Time (s) | Testing Time (s) | SI Accuracy (%) | SV EER (%) |
|---|---|---|---|---|---|
| Fusion (MFCC+LPC+PLP+ +entropy+RMS+centroid)+ICA | KNN | 0.001 | 0.012 | 96.62 | 1.72 |
| ΔFusion+ICA | KNN | 0.001 | 0.015 | 93.24 | 1.19 |
| ΔΔFusion+ICA | KNN | 0.001 | 0.012 | 94.23 | 1.01 |
| Fusion+ΔFusion+ICA | KNN | 0.001 | 0.015 | 95.82 | 1.28 |
| Fusion+ΔΔFusion+ICA | KNN | 0.001 | 0.013 | 97.01 | 0.98 |
| ΔFusion+ΔΔFusion+ICA | KNN | 0.001 | 0.013 | 95.02 | 0.89 |
| Fusion+ΔFusion+ΔΔFusion+ICA | KNN | 0.001 | 0.011 | 96.42 | 0.79 |

Tables 3-6 present the calculated accuracy of speaker identification (SI) and the equal error rate (EER) for PCA and ICA methods in reducing the highest-dimensional feature space. According to these results, the PCA method yielded the best outcomes. When using the LD (linear discriminant) classification method, the best SI accuracy was 98.6% with an EER of 0.22%. Additionally, when employing the KNN (K-nearest neighbor) classifier, the best SI accuracy was recorded at 97.01% with an EER of 0.79%.

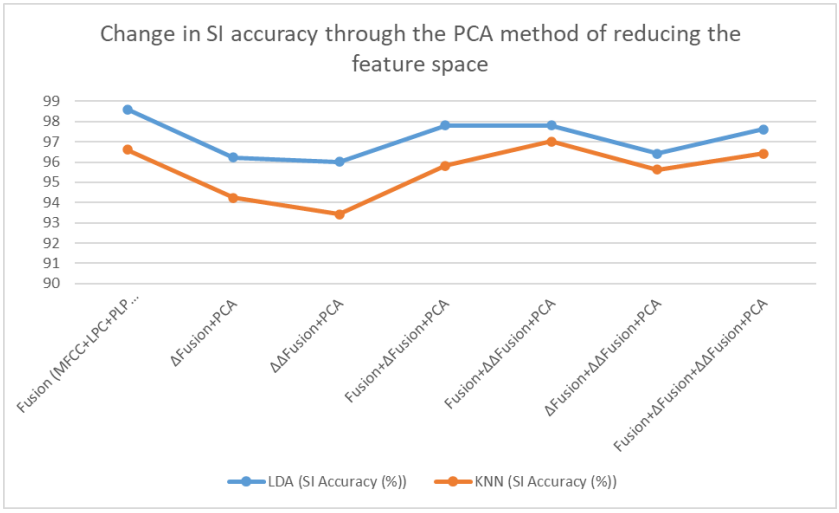



**Fig. 6.** Graph showing the change in SI accuracy through the PCA method of reducing the feature space.

The best results obtained using the ICA method were achieved when applying the LD (linear discriminant) classification method, with the highest SI accuracy of 98.4% and an equal error rate (EER) of 0.22%. Additionally, when using the KNN (K-nearest neighbor) classifier, the best SI accuracy was 96.62% with an EER of 0.79%.

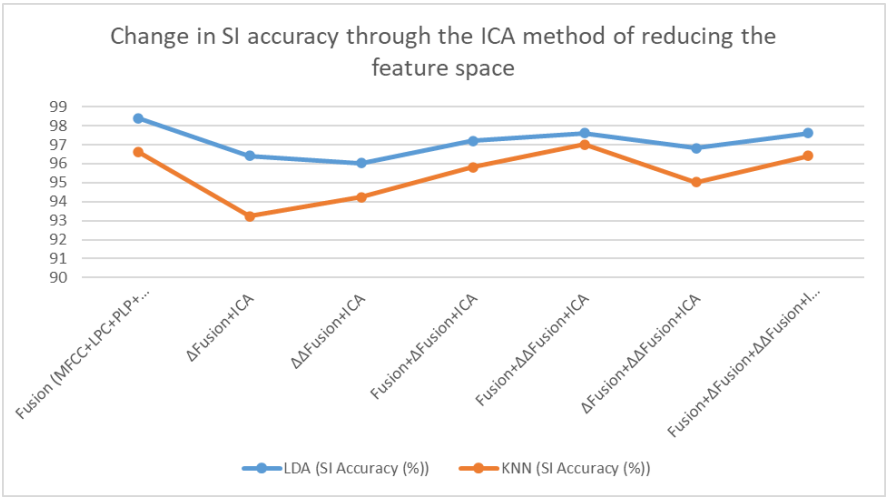



**Fig. 7.** Graph showing the change in SI accuracy through the ICA method of reducing the feature space

### 5.3 System Configuration

The research used PYTHON software and NVIDIA GeForce RTX 3090 Ti (24GB) GPUs for calculating training and testing times.

## 6 Conclusions

In this research, we attempted to solve the problem of speaker recognition through less informative learning through various approaches. We proposed a novel feature extraction method based on feature space reduction techniques. Our research encompassed an in-depth study of speaker recognition, feature fusion, and the application of Principal Component Analysis (PCA) and Independent Component Analysis (ICA) methods for feature space reduction. Our newly proposed feature space reduction method, applied to reduced feature vectors, partially improved speaker recognition performance across various classification methods. Notably, we increased speaker recognition accuracy to 97.4% through feature fusion. Additionally, when applying PCA and ICA feature space reduction methods along with LD and KNN classifiers, we achieved an Equal Error Rate (EER) of 0.22% in speaker verification. In conclusion, the LDA classifier demonstrated high effectiveness when used in combination with feature space reduction methods across various noise levels and dataset sizes, indicating its potential for practical applications.

## References

1. Chauhan, N., Isshiki, T., Li, D.: Enhancing Speaker Recognition Models with Noise-Resilient Feature Optimization Strategies. Acoustics 2024, 6(2), 439-469.
2. Mirzaev, N., Urinboev, J., Gafforov, N., Nugmanova, M.: Analysis of Speech Signal Processing Methods in Speech Recognition Systems. IEEE 6th International Conference on Problems of Cybernetics and Informatics (PCI), Baku, Azerbaijan, 25–28 august 2025, pp. 1-8,
3. Zamalloa, M., Bordel, G., Rodriguez, L., Penagarikano, M.: Feature selection based on genetic algorithms for speaker recognition. In Proceedings of the 2006 IEEE Odyssey—The Speaker and Language Recognition Workshop, San Juan, PR, USA, 28–30 June 2006, pp. 1–8.
4. Furui, S.: Comparison of speaker recognition methods using statistical features and dynamic features. IEEE Trans. Acoust. Speech Signal Process. 1981, 29, 342–350.
5. Hermansky, H., Morgan, N., Bayya, A., Kohn, P.: Compensation for the effect of the communication channel in auditory-like analysis of speech (RASTA-PLP). In Proceedings of the 2nd European Conference on Speech Communication and Technology (Eurospeech 1991), Genovo, Italy, 24–26 September 1991, pp. 1367–1370.
6. Adami, A.G., Mihaescu, R., Reynolds, D.A., Godfrey, J.J.: Modeling prosodic dynamics for speaker recognition. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, 6–10 April 2003, pp. IV–788.
7. Sönmez, K., Shriberg, E., Heck, L., Weintraub, M.: Modeling dynamic prosodic variation for speaker verification. In Proceedings of the 5th International Conference on Spoken

Language Processing (ICSLP 1998), Sydney, Australia, 30 November–4 December 1998, pp. 3189–9192.
8. Kumar, K., Kim, C., Stern, R.M.: Delta-spectral cepstral coefficients for robust speech recognition. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011, pp. 4784–4787.
9. Carey, M.J., Parris, E.S., Lloyd-Thomas, H., Bennett, S.: Robust prosodic features for speaker identification. In Proceedings of the Fourth International Conference on Spoken Language Processing, ICSLP 9‘6, Philadelphia, PA, USA, 3–6 October 1996, pp. 1800–1803.
10. Chauhan, N., Isshiki, T., Li, D.: Speaker recognition using LPC, MFCC, ZCR features with ANN and SVM classifier for large input database. In Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 23–25 February 2019, pp. 130–133.
11. Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. 2011, 19, 788–798.
12. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process. 1980, 28, 357–366.
13. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons: New York, NY, USA, (2001).
14. Rosca, J., Kopfmehl, A.: Cepstrum-like ICA representations for text independent speaker recognition. In Proceedings of the ICA’2003, Nara, Japan, 1–4 April 2003, pp. 999–1004.
15. Ding, P., Kang, X., Zhang, L.: Personal recognition using ICA. In Proceedings of the ICONIP (2001).
16. Cichocki, A., Amari, S.I.: Adaptive Blind Signal and Image Processing. John Wiley: Chichester, UK, (2002).
17. Mirzaev, N., Urinboyev, J.: Methods of Feature Extraction in Speaker Recognition and Their Effectiveness. International Scientific and Technical Conference Current State and Prospects for the Development of Digital Technologies and Artificial Intelligence, Bukhara, September 27-28, 2024, pp.102-108.
18. Mirzaev, N., Urinboyev, J, Nugmanova, M.: Main Methods of Speaker Identification Through Speech. Journal of Transport 2024, 1(4), 130-136.
19. Mirzaev, N., Urinboyev, J.: Methods of linear prediction and spectral centroid in the selection of features characterizing sound in the speaker recognition problem. Digital transformation and artificial intelligence: problems, innovations and trends 1st international scientific - practical conference, TSUE, Tashkent, September 11, 2024, pp. 29-30.
20. Mirzayev, N., Urinboyev, J.: Mel-frequency cepstral analysis method for extracting voice characteristics in speaker recognition. Computational models and technologies (CMT2024) Third International conference dedicated to the 90th birth anniversary of Professor M.I. Isroilov, Tashkent, April 24, 2024, pp. 199-201.
21. Chakroborty, S., Roy, A., Saha, G.: Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification. In Proceedings of the 2006 IEEE International Conference on Industrial Technology, Mumbai, India, 15–17 December 2006, pp. 387–390.
22. Chauhan, N., Isshiki, T., Li, D.: Speaker Recognition using fusion of features with Feedforward Artificial Neural Network and Support Vector Machine. In Proceedings of the 2020 international conference on intelligent engineering and management (ICIEM), London, UK, 17–19 June 2020, pp. 170–176.

23. Ahmad, K.S., Thosar, A.S., Nirmal, J.H., Pande, V.S.: A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network. In Proceedings of the 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, India, 4–7 January 2015, pp. 1–6.
24. Wang, L., Chen, Z., Yin, F.: A novel hierarchical decomposition vector quantization method for high-order LPC parameters. IEEE/ACM Trans. Audio Speech Lang. Process. 2014, 23, 212–221.
25. Slifka, J., Anderson, T.R.: Speaker modification with LPC pole analysis. In Proceedings of the 1995 International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA, 9–12 May 1995, pp. 644–647.
26. Mirzaev, O, Radjabov, S, Mirzaev, N, Meliev, F, Tillavoldiev, A.: A family of recognition algorithms based on the construction of k-dimensional threshold rules. Procedia Computer Science, 2024 (237), 610-617.
27. Chauhan, N., Chandra, M.: Speaker recognition and verification using artificial neural network. In Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 22–24 March 2017, pp. 1147–1149.
28. Hermansky, H.: Perceptual linear predictive (PLP) analysis of speech. J. Acoust. Soc. Am. 1990 (87), 1738–1752.
29. Root-mean-square Value.: A Dictionary of Physics, 6th ed., Oxford University Press: Oxford, UK, (2009).
30. Ross, A.: Fusion, feature-level. In Encyclopedia of Biometrics, Li, S.Z., Jain, A., Eds., Springer: Boston, MA, USA, 2009, 597–602.
31. You, S.D., Hung, M.J.: Comparative study of dimensionality reduction techniques for spectral–temporal data. Information 2021, 12(1), 1-12.
32. Herault, J., Jutten, C., Ans, B.: Detection de grandeurs primitives dans un message composite par une architecture de calcul neuromimetique en apprentissage non supervise. In Proceedings of the GRETSI, Nice, France, 20–24 May 1985, pp. 536.
33. Opanasenko, V., Fazilov, Sh., Mirzaev, N., Kakharov, Sh.: A Model of Recognition Algorithms Based on Threshold Functions and Object Proximity Assessment. Cybernetics and Systems Analysis, 61(1), (2025), 148-160.
34. Tharwat, A.: Independent component analysis: An introduction. Appl. Comput. Inform. 2018, 17, 222–249.
35. Yao, Z., Ruzzo, W.L.: A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data. BMC Bioinform. 2006, 7, S11. 1-11
36. Subasi, A.: Machine learning techniques. In Practical Machine Learning for Data Analysis Using Python, Elsevier: Amsterdam, The Netherlands, 2020, pp. 91–202.
37. Tharwat, A.: Classification assessment methods: A detailed tutorial. Appl. Comput. Inform. 2020, 17, 168–192.
38. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 2015, 10 (e0118432), 1-21.

# Mapping Mobility Justice: A Deep Learning and GIS-Based Framework for Micro-Scale Sustainable Mobility Policy Assessment

Süleyman Nurullah Adahi ŞAHİN[1] [0000-0002-9232-8104], Abdulkadir ÖZDEN[1] [0000-0002-3657-890X] Hakan KOCAMAN[2] [0000-0002-0986-1339], Abdurrahman KORKMAZ[3] [0000-0002-0684-2941] , Yavuz HALİMERGÜN[2] [0000-0001-7363-4029] and Cem ÖZKURT[4] [0000-0002-1251-7715]

[1] Civil Engineering Department, Sakarya University of Applied Sciences, Sakarya, Türkiye
[2] Postgraduate Education Institute, Sakarya University of Applied Sciences, Sakarya, Türkiye
[3]Keşap Vocational School, Giresun University, Keşap, Giresun, Türkiye
[4]Computer Engineering Department, Sakarya University of Applied Sciences, Sakarya, Türkiye
lncs@springer.com

**Abstract.** The development of inclusive and sustainable transport policies demands accurate, spatially detailed data on the built environment and mobility infrastructure—particularly at the neighborhood level, where daily mobility patterns and access disparities are most visible. This study proposes MiSHar (Micro-Scale Sustainable Mobility Index), an integrated analytical framework that leverages deep learning-based object detection and Geographic Information Systems (GIS) to systematically evaluate the sustainable mobility potential of urban micro-regions. High-resolution satellite and aerial imagery are processed using convolutional neural networks to automatically detect physical features such as buildings, green areas, paved surfaces, pedestrian and bicycle networks. These are subsequently combined with GIS-based public transport bus stop yielding a multi-dimensional index formulation. The index incorporates weighted parameters related to population density, infrastructure provision, and land use patterns. A pilot implementation in Adapazarı District (Sakarya, Türkiye) demonstrates the model's ability to reveal intra-urban mobility inequalities and provide spatial evidence for transport-oriented development strategies. MiSHar represents a policy-responsive, survey-independent, and scalable approach to mobility diagnostics. By integrating AI-powered spatial analytics with urban planning principles, it offers a valuable tool for local governments seeking to target infrastructure investments, design equity-focused interventions, and align with European and global sustainable mobility goals.

**Keywords:** Sustainable mobility index, Deep learning, GIS, Micro-scale analysis

## 1 First Section

Urban mobility systems are undergoing profound transformations as cities grapple with rapid population growth, rising vehicle ownership, and escalating demands for equitable access to opportunities. In Europe and globally, sustainable mobility has

been positioned as a cornerstone of climate resilience and social inclusion, reflected in the European Green Deal, Urban Mobility Framework (EC, 2021), and the Sustainable Development Goals (SDG 11: Sustainable Cities and Communities). Yet, achieving these goals requires more than macro-scale transport planning; it necessitates granular, micro-scale analyses of how mobility infrastructures and services are distributed across neighborhoods.

Traditional transport planning models—such as the four-step travel demand framework—have proven valuable for forecasting aggregate travel behavior. However, their effectiveness is highly dependent on the definition of spatial units. In many developing countries, including Türkiye, traffic analysis zones (TAZs) are defined at relatively large scales, often encompassing entire neighborhoods or even broader areas (Altan & Ayözen, 2018; McNally, 2007). While this coarse zoning structure reduces the burden of data collection and survey design, it also masks fine-grained variations in mobility patterns. As a result, although intra-urban disparities remain less visible, the use of large-scale TAZs in Türkiye contributes to strengthening the spatial coverage and applicability of the traditional model. Consequently, the equity and sustainability dimensions of daily mobility—particularly walkability, cycling potential, and access to public transport—remain underexplored at the neighborhood level.

Emerging advances in GIS and artificial intelligence (AI) provide opportunities to overcome these limitations. Deep learning techniques applied to high-resolution satellite imagery enable automated extraction of built environment features, while open datasets such as OpenStreetMap (OSM) and municipal transit records allow integration of infrastructure and service attributes. Together, these tools make it possible to construct fine-grained indices of sustainable mobility potential without the prohibitive costs of large-scale surveys (Liu et al., 2020; Esch, 2011).

This paper introduces MiSHar a framework designed to quantify sustainable mobility capacity at the sub-neighborhood scale. By combining deep learning–based building and land-use detection with GIS-based transit and active transport infrastructure data, MiSHar provides a composite measure of mobility justice. A pilot study is conducted in Karaman/Yenikent region of Sakarya (Türkiye), where rapid urbanization, mixed land-use patterns, and increasing travel demand create both opportunities and challenges for sustainable mobility.

The contributions of this study are threefold:

1. Methodological innovation – integrating AI-powered remote sensing with GIS to produce a scalable, survey-independent mobility index.
2. Policy relevance – generating actionable spatial evidence to support transport-
3. Empirical application – demonstrating the feasibility of MiSHar through a pilot implementation in a rapidly growing Turkish urban district.

By situating mobility analysis at the micro-scale, this research aligns with TRA 2026's theme of "Advancing Sustainable and Inclusive Mobility", providing a replicable framework for European and global cities seeking to bridge the gap between smart data analytics and socially just transport policies.

## 2 Literature Review

Sustainable urban mobility has long been recognized as a multi-dimensional challenge encompassing environmental, economic, and social objectives (Litman & Burwell, 2006). The European level emphasize integrated, people-centered planning approaches. Internationally, frameworks such as the UN Sustainable Development Goals and the Paris Agreement further highlight mobility's role in reducing carbon emissions and promoting equitable access.

Despite this policy momentum, the translation of strategic goals into localized interventions remains difficult. Scholars stress that accessibility, rather than mobility alone, should be the central metric for transport justice (Boschmann & Kwan, 2008). Accessibility encompasses not only the physical presence of infrastructure but also its affordability, safety, and usability across diverse social groups.

While metropolitan-scale transport models dominate practice, micro-scale assessments provide crucial insights into neighborhood-level disparities. Studies on walkability indices (Chatziioannou et al., 2023), school-area accessibility metrics (Prado, 2016), and sustainable mobility scales (Saiu et al., 2022) highlight how localized indicators can reveal inequities hidden in aggregated data. However, many of these tools rely on manual surveys or limited datasets, constraining their applicability across diverse urban contexts.

In Türkiye, reliance on traditional TAZ structures has further limited precision. Household surveys underpinning Istanbul's master plans in 1997 and 2012 grew from 12,000 to 160,000 responses, yet still struggled to capture heterogeneity at the sub-neighborhood scale (Altan & Ayözen, 2018). For mid-sized cities such as Sakarya, comprehensive survey efforts are often absent, creating a critical data gap in planning.

GIS has become an indispensable tool for transport and urban planning, enabling integration of land use, demographic, and infrastructure data. Remote sensing adds a powerful dimension by providing consistent, high-resolution imagery for built environment monitoring. Applications range from building density estimation (Kajimoto & Susaki, 2013; Xu et al., 2022) to green space mapping (Liu et al., 2020). Such data sources can support fine-grained accessibility analyses, particularly when traditional survey data are lacking. Attempts to operationalize sustainability at the micro-scale have produced various composite indices. For example, ElDessouki (2022) proposed a time-based equity index, while the IMPES model (Prado, 2016) evaluated pedestrian accessibility around schools. Yet most indices remain limited in scope, focusing on single modes or specific contexts.

The MiSHar framework builds on these precedents by combining multiple dimensions—built environment density, transport infrastructure, population distribution, and green space availability—into a unified index. Unlike survey-heavy approaches, MiSHar relies on open-source datasets (OSM), municipal transit records, and AI-based image analysis. This not only reduces costs but also enhances scalability, making it suitable for mid-sized cities where resources are limited. Existing literature demonstrates the potential of GIS and deep learning for mobility assessment but stops short of integrating these methods into comprehensive policy-oriented tools. Moreo-

ver, few studies have applied such approaches in the Turkish context, where urban growth pressures are intense and planning data are sparse.

This study addresses these gaps by:

- Developing a survey-independent mobility index tailored to micro-scale analysis,
- Demonstrating its application in Karaman/Yenikent, Sakarya, and
- Positioning the results within broader European sustainable mobility policy debates.

## 3 Methodology

This study develops the Micro-Scale Sustainable Mobility Index (MiSHar) for the Karaman/Yenikent district of Sakarya by combining Geographic Information Systems (GIS), deep learning–based feature extraction, and statistical weighting approaches. The index is designed to provide a standardized and reproducible measure of sustainable mobility capacity at the micro-zone level, avoiding the limitations of survey-dependent transport planning methods

These spatial features were subsequently integrated with open data from OpenStreetMap (OSM) and municipal datasets, which provided information on pedestrian and cycling networks, public transport stops, and service coverage. The combined dataset allowed for the computation of six key indicators (Table 1):

- Green area (GA) – total green space in each zone,
- Building density (BD) – ratio of built-up area to total area,
- Pedestrian network length (PL),
- Cycling network length (CL),
- Public transport stops (PTS),
- Population density (PD).

**Table 1.** Pilot Study Area Data extracted from OSM database

| Zone | Green Area ($m^2$) | Number of Building | Building Density | Pedestrian Road Distance (m) | Cycling Path Distance (m) | Number of Public Transport Stop | Population Density (person/$m^2$) |
|---|---|---|---|---|---|---|---|
| 1 | 300,96 | 61 | 0,0999 | 5660,44 | 0 | 7 | 0,012887734 |
| 2 | 300,96 | 75 | 0,1107 | 7697,67 | 0 | 14 | 0,014066377 |
| 3 | 723,41 | 76 | 0,0696 | 13533,36 | 1361,4 | 18 | 0,006486118 |
| 4 | 20396,250 | 53 | 0,0808 | 3464,22 | 0 | 8 | 0,008541925 |
| 5 | 0 | 83 | 0,12 | 8138,56 | 0 | 11 | 0,01070640 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | 18 | | | | 4 |
| 6 | 129 60,14 | 49 | 0,14 10 | 6747,47 | 0 | 7 | 0,00815508 9 |
| 7 | 170 98,75 | 94 | 0,14 45 | 8185,84 | 345,67 | 12 | 0,01126761 5 |
| 8 | 951 4,07 | 47 | 0,09 07 | 6031,89 | 1015,73 | 9 | 0,00829556 7 |
| 9 | 499 9,24 | 43 | 0,11 50 | 4395,69 | 0 | 3 | 0,01212293 8 |

These variables were normalized by dividing by their observed maxima across the study area, used here as proxies for theoretical maxima. The general formulation of MiSHar is:

$$MiSHar = (\alpha \cdot (GA/GAmax) + \beta \cdot (BD/BDmax) + \gamma 1 \cdot (PL/PLmax) + \gamma 2 \cdot (CL/CLmax) + \gamma 3 \cdot (PTS/PTSmax) + \delta \cdot (PD/PDmax)) \quad (1)$$

Three different approaches were applied to determine these weights:

- Equal weights: each component assigned the same importance (1/6). This baseline approach assumes equal contribution of all factors to sustainable mobility.
- Policy weights: higher priority was assigned to active and public transport components (PL, BD, PTS each weighted 0.20), moderate weight to green space and building density (0.15 each), and lower weight to population density (0.10). This reflects policy priorities in European sustainable mobility strategies, where accessibility and low-carbon modes are central.
- PCA-derived weights: to avoid subjective assumptions, a Principal Component Analysis (PCA) was conducted on the normalized variables. The loadings from the first principal component (explaining ~50% of total variance) were converted into positive, normalized weights summing to one. This unsupervised method objectively identified the relative importance of each component in explaining variation in mobility-related attributes.

The MiSHar index was then computed for each micro-zone under all three weighting schemes, producing values on a 0–1 scale. While equal and policy-weighted indices provide useful benchmarks, the PCA-based index was preferred as the primary analytical tool in this study because it reflects the inherent structure of the data without requiring expert judgment or external calibration.

## 4 Results

The MiSHar index was calculated for nine micro-zones in the Karaman/Yenikent district (Figure 1), applying the three weighting schemes defined in the methodology: equal weighting, policy weighting, and PCA-derived weighting. Across all approaches, MiSHar values were normalized on a 0–1 scale, with higher scores reflecting greater sustainable mobility potential.

**Fig. 1.** Study Area (Karaman/Yenikent, Sakarya)

The equal-weighted MiSHar produced a relatively compressed distribution (mean = 0.390, SD = 0.055), ranging from 0.303 to 0.488. This smoothing effect arises because all factors contribute equally, so weaknesses in one dimension are balanced by strengths in another. In contrast, the policy-weighted MiSHar introduced greater differentiation (mean = 0.351, SD = 0.067; range = 0.260–0.460), with higher emphasis placed on zones with robust transit and active mobility infrastructure (Figure 2)

The PCA-based MiSHar generated the most pronounced contrasts, with scores spanning from 0.240 to 0.510. Because PCA empirically derives weights from the dataset, it amplified disparities that were less visible under equal and policy weighting. Zones combining dense pedestrian and cycling networks with multiple public transport stops consistently achieved the highest scores, while peripheral areas lacking such infrastructure were downgraded.

The PCA index results reveal a distinct intra-district hierarchy. Centralized zones with greater connectivity and multimodal service provision scored highest, underscor-

ing the decisive role of functional accessibility infrastructure. Conversely, areas dominated by green space or moderate building density but lacking public transport and active mode networks exhibited weaker sustainable mobility performance. This highlights that infrastructure enabling access—rather than simply land-use form—constitutes the critical determinant of mobility justice at the neighborhood scale.

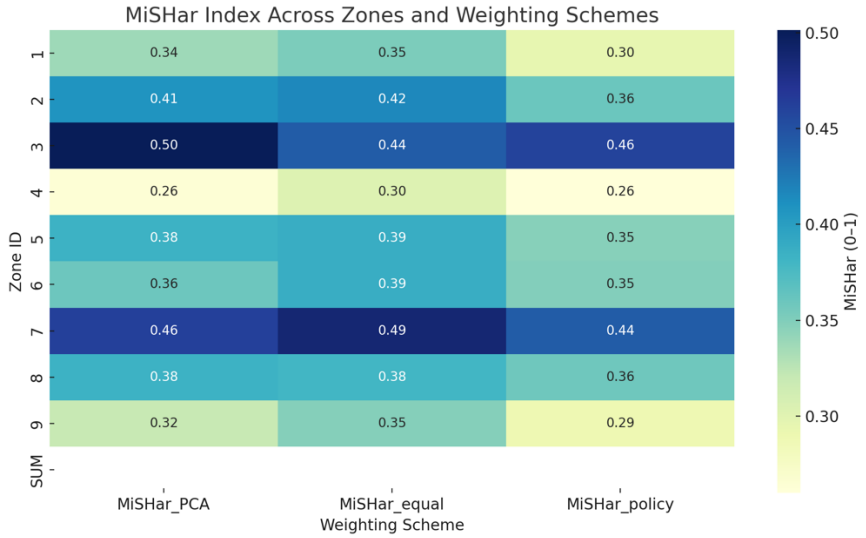



**Fig 2**. MiSHar Index Across Zones

The PCA model explained approximately 50% of the total variance in the six input dimensions. Factor loadings indicated that public transport stops (PTS), pedestrian network length (PL), and cycling network length (CL) carried the greatest weights, jointly accounting for nearly two-thirds of the PCA variance. Green area (GA) and population density (PD) received lower weights, reflecting their weaker contribution to differentiating micro-zones. These findings are consistent with international evidence that accessibility and multimodality are stronger predictors of mobility sustainability than morphological indicators alone (Boschmann & Kwan, 2008; Saiu et al., 2022).

## 5 Discussion and Conclusion

The results of this study demonstrate that the design of weighting schemes has a decisive influence on how sustainable mobility capacity is measured and interpreted at the micro-scale. The equal-weight formulation provided a neutral benchmark, producing

balanced yet relatively homogenous scores. While useful as a diagnostic baseline, this approach risks obscuring the differentiated role of transport infrastructure in shaping accessibility. The policy-weighted index, by contrast, aligned directly with European mobility agendas such as the Urban Mobility Framework and the European Green Deal, rewarding zones with stronger provision of active and public transport facilities. Yet this method presupposes normative priorities, which may not fully reflect the complex realities of diverse urban contexts.

The PCA-derived MiSHar index proved most robust by empirically estimating weights from the dataset itself. This data-driven approach minimized subjectivity and revealed that public transport stops, pedestrian pathways, and cycling networks are the dominant contributors to intra-district variations in sustainable mobility. By contrast, morphological attributes such as green space and building density—though valuable for broader urban quality of life—played a secondary role in determining accessibility and mobility equity. The finding that PCA explained approximately half of the total variance across all dimensions underscores the methodological strength of allowing the data, rather than expert judgment, to determine factor importance.

From a policy perspective, the implications are clear: investments in functional accessibility infrastructure yield disproportionately greater improvements in sustainable mobility capacity than marginal increases in green area or adjustments to building density. For municipalities such as Sakarya, this provides a powerful, evidence-based rationale for prioritizing the expansion of pedestrian and cycling networks and the densification of public transport stops. By shifting emphasis from form to function, MiSHar supports the evolution of planning paradigms from mobility to accessibility, ensuring that investments directly enhance equity, usability, and climate resilience.

In conclusion, MiSHar offers a scalable, transferable, and survey-independent framework for diagnosing and addressing intra-urban mobility disparities. By integrating AI-powered feature detection, GIS-based analysis, and statistically derived weighting, it bridges methodological innovation with policy relevance. Applied to Karaman/Yenikent, the framework not only illuminated local inequalities but also provided actionable guidance that aligns with European climate and mobility objectives. In doing so, MiSHar exemplifies the advancing sustainable and inclusive mobility through approaches that are rigorous, data-driven, and policy-responsive.

**References**

Altan, M., & Ayözen, A. (2018). Ulaşım ana planları ve hanehalkı anketleri üzerine değerlendirmeler: İstanbul örneği. İstanbul Büyükşehir Belediyesi Ulaşım Dairesi Yayınları.

Arıkan, D., & Yıldız, F. (2023). Göktürk-1 uydu görüntülerinden U-Net modeli kullanılarak binaların çıkarılması. Turkish Journal of Remote Sensing. https://doi.org/10.51489/tuzal.1300939

Bachl, B. (2019). CityGAN: Learning urban planning with generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1–8.

Beber, M. (2021). Walkability indices and urban mobility: Toward sustainable accessibility. Journal of Urban Planning and Development, 147(2), 04021009.

Bellini, F., Bolognesi, M., Nesi, P., Pantaleo, G., & Paolucci, M. (2019). SUITS: Supporting small and medium-sized cities in sustainable urban mobility transition. Transportation Research Procedia, 37, 177–184. https://doi.org/10.1016/j.trpro.2018.12.178

Boschmann, E. E., & Kwan, M. P. (2008). Toward socially sustainable urban transportation: Progress and potentials. International Journal of Sustainable Transportation, 2(3), 138–157. https://doi.org/10.1080/15568310701517265

Chatziioannou, I., Guth, D., & Hillel, T. (2023). Classifying urban walkability through multi-dimensional indicators. Transportation Research Part D: Transport and Environment, 116, 103639. https://doi.org/10.1016/j.trd.2023.103639

ElDessouki, A. (2022). Time-based accessibility equity index: An application for sustainable urban mobility. Journal of Transport Geography, 102, 103376. https://doi.org/10.1016/j.jtrangeo.2022.103376

Esch, T., et al. (2011). Urban footprint processor—Fully automated processing chain generating settlement masks from medium-resolution satellite data. Remote Sensing of Environment, 115(9), 2347–2360. https://doi.org/10.1016/j.rse.2011.04.027

Kajimoto, M., & Susaki, J. (2013). Estimation of building density using polarimetric synthetic aperture radar images. IEEE Transactions on Geoscience and Remote Sensing, 51(3), 1793–1804. https://doi.org/10.1109/TGRS.2012.2207127

Kocaman, H., Özden, A., Şahin, S. N. A., Özkurt, C., & Korkmaz, A. (2025). Deep learning-based building detection for micro-scale transportation planning: GIS and ITS applications. In Proceedings of the 1st National Intelligent Transportation Systems Conference (Conf-ITS'25) (pp. 1–10).

Litman, T., & Burwell, D. (2006). Issues in sustainable transportation. International Journal of Global Environmental Issues, 6(4), 331–347. https://doi.org/10.1504/IJGENVI.2006.010889

Liu, C., Hu, Y., & Liu, Y. (2020). Building density estimation from high-resolution aerial imagery using machine learning techniques. Remote Sensing, 12(18), 2967. https://doi.org/10.3390/rs12182967

McNally, M. G. (2007). The four-step model. In D. A. Hensher & K. J. Button (Eds.), Handbook of Transport Modelling (2nd ed., pp. 35–52). Emerald Group Publishing.

Miller, H. J. (2021). The problem of transport analysis zones (TAZ) in accessibility studies. Journal of Transport Geography, 93, 103084. https://doi.org/10.1016/j.jtrangeo.2021.103084

Mouzas, S. (2022). Evaluating sustainable mobility policies: A conceptual framework. Sustainability, 14(7), 4125. https://doi.org/10.3390/su14074125

Naito, S., et al. (2020). Building-damage detection method based on machine learning utilizing aerial photographs of the Kumamoto earthquake. Earthquake Spectra, 36(3), 1166–1187. https://doi.org/10.1177/8755293019901309

Prado, V. (2016). Microaccessibility index in surrounding school areas: Measuring pedestrian safety and access equity. Transportation Research Procedia, 18, 36–43. https://doi.org/10.1016/j.trpro.2016.12.006

Rakodi, C. (2001). Forget planning, put politics first? Priorities for urban management in developing countries. International Journal of Applied Earth Observation and Geoinformation, 3(3), 209–223. https://doi.org/10.1016/S0303-2434(01)85026-X
Saiu, V., et al. (2022). Measuring sustainable mobility performance of neighborhoods: An index-based approach. Sustainability, 14(12), 7339. https://doi.org/10.3390/su14127339
Xu, Y., Wu, C., & Zhang, J. (2022). Estimating building height and density from street-view imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 183, 236–250. https://doi.org/10.1016/j.isprsjprs.2021.10.019

# A Comparative Study of Word Embedding Techniques for Turkish Sentiment Analysis

Buse Sarıçayır[1][0009-0005-5868-0189], Caner Özcan [0000-0002-2854-4005], and Yasin Ortakçı [0000-0002-0683-2049]

[1] Karabuk University, Karabuk 78000, Türkiye
busesaricayir@gmail.com

**Abstract.** This paper presents a comparative analysis of three different word embedding techniques—TF-IDF, Word2Vec, and FastText in the context of Turkish sentiment analysis. Each method offers a unique trade-off between computational efficiency, representational richness, and classification accuracy of Turkish text. TF-IDF, a frequency-based approach, is computationally inexpensive but neglects word order and contextual information. Word2Vec generates context-aware embeddings capturing semantic relationships but remains context-independent for individual words. FastText further refines word representation by incorporating subword information, which is particularly advantageous for morphologically rich languages like Turkish. The performance of these embedding methods was evaluated using four different classification models (Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine) on the Winvoker Turkish sentiment analysis dataset. Our experimental results show that Word2Vec achieves the highest accuracy with Logistic Regression. These results contribute to a better understanding of the strengths and limitations of various word embedding techniques for sentiment analysis in morphologically rich languages.

**Keywords:** Sentiment Analysis, Turkish text, TF-IDF, Word2Vec, FastText, Embedding Models.

## 1 Introduction

Sentiment analysis, the computational process of identifying and categorizing opinions expressed in text, has become an indispensable tool for understanding human behavior and extracting valuable insights from the ever-increasing volume of unstructured data generated daily. Its applications are widespread, ranging from social media monitoring and brand reputation management to customer feedback analysis, market research, and political science [1]. The accuracy and effectiveness of sentiment analysis systems depend heavily on the ability to effectively represent the meaning and context of words within sentences and documents. This representation significantly affects the classifier's ability to detect subtle nuances in sentiment expressions, such as sarcasm, irony, and negation [2].

This study focuses on sentiment analysis of Turkish text, a particularly challenging domain due to the complexities of the Turkish language, including its agglutinative nature and rich morphology. The agglutinative structure, where multiple suffixes are

attached to a root word, requires sophisticated linguistic processing to accurately capture the intended meaning and sentiment. Furthermore, the prevalence of informal language and colloquialisms in online Turkish text presents additional obstacles for accurate sentiment classification. Addressing these challenges is crucial because effective sentiment analysis of Turkish text offers significant opportunities, including deeper understanding of public opinion in Turkey, improved market research for Turkish businesses, and enhanced customer service experiences for Turkish-speaking consumers. This study aims to contribute to the advancement of sentiment analysis techniques specifically tailored to the nuances of the Turkish language.

We conduct a comparative analysis of three word embedding methods—TF-IDF, Word2Vec, and FastText—for the task of Turkish sentiment analysis. TF-IDF, known for its computational efficiency and ease of implementation, represents words as vectors based solely on their frequency within a document and across the corpus [3]. However, it lacks the ability to capture contextual information and word order, both of which are critical for accurately conveying sentiment [4]. Word2Vec [5], on the other hand, generates context-aware embeddings that capture semantic relationships between words through their co-occurrence patterns [6]. Despite this advantage, Word2Vec treats each word independently of its context, which limits its ability to distinguish between multiple meanings of a word [7]. FastText addresses this limitation by incorporating subword information, such as n-grams, which is particularly advantageous for handling the morphological complexity of the Turkish language and alleviating challenges associated with rare or out-of-vocabulary words [8]. By providing a more nuanced representation of sentiment, FastText offers a potential improvement over the other methods. The analysis seeks to identify the most effective embedding technique for Turkish sentiment analysis [9].

This paper presents an analysis of three word embedding techniques for Turkish sentiment analysis, using the Winvoker dataset [10] as our corpus. Our experiments, conducted using the Winvoker Turkish sentiment dataset and four machine learning models, demonstrate that Word2Vec consistently outperforms TF-IDF and achieves accuracy comparable to FastText. These results highlight the importance of considering contextual information and subword morphology for effective sentiment analysis in morphologically rich languages like Turkish. A detailed analysis of the results, including confusion matrices and performance metrics for each model and embedding technique, is presented in the following sections.

## 2 Related Work

Recent research on word embedding methods for Turkish sentiment analysis reveals a trend towards increasingly sophisticated deep learning approaches, although simpler methods still have value in certain contexts. Early work relied on traditional machine learning techniques like Support Vector Machines (SVM) and Logistic Regression (LR). Demircan et al. [11], using Turkish e-commerce reviews, achieved an accuracy of 87% and an F1-score of 0.86 using SVM and feature selection, highlighting the importance of careful feature engineering even with simpler models. Similarly, Aydın

and Gungor [12] found that supervised techniques such as SVM and Random Forest (RF) outperformed semi-supervised and unsupervised methods (88% accuracy versus 70% and 65%, respectively) on various datasets including reviews and news articles. Kemaloglu et al. [13], working with noisy social media data, obtained an accuracy of 84% and F1-score of 0.82 using logistic regression with TF-IDF, showcasing the challenges inherent in such datasets. Ozdemir and Ortakci [14] further demonstrate the continued relevance of frequency-based methods by achieving 84% accuracy using TF-IDF on a call center dataset.

The introduction of transformer-based models has significantly improved performance. Acikalin et al. [15] achieved 90% accuracy and 0.89 F1-score using BERT on a custom Turkish sentiment dataset, outperforming traditional methods by 15%. Guven [16], using BERTurk with a novel text filtering technique on benchmark datasets, further enhanced accuracy to 92.5% and F1-score to 0.93, underscoring the impact of data preprocessing. Ba Alawi & Bozkurt [17] corroborated these findings, demonstrating that BERTurk outperformed traditional word embeddings by 10% in accuracy (91%) and F1-score (0.90) across diverse datasets. Alshammari & Akyüz [18] used a RoBERTa model fine-tuned for Turkish tweets to achieve an impressive F1-score of 0.92 in analyzing mental health-related sentiments. However, Cam et al. [19] showed that with careful feature engineering, Gradient Boosting could still achieve high accuracy (89%) and F1-score (0.88) on financial sentiment tweets.

Automation of model training is also promising. Tasar et al. [20] utilized AutoTrain to achieve 86% accuracy and 0.84 F1-score on various Turkish sentiment datasets, suggesting a streamlined approach for achieving near state-of-the-art results. Finally, Ozdemir and Ortakci [21] combined hybrid vectorization with deep learning models for a call center routing task, obtaining 85% accuracy and an F1-score of 0.83, highlighting the potential of tailored approaches for specific applications. Overall, these studies indicate that while sophisticated deep learning models currently dominate, simpler methods combined with effective feature engineering can still be competitive, especially in specialized domains with limited data or specific requirements.

While transformer-based models such as BERTurk and mBERT have demonstrated superior performance in Turkish sentiment analysis, this study intentionally focuses on classical embedding methods to provide a controlled comparison of TF-IDF, Word2Vec, and FastText. Transformer-based approaches are left for future work.

## 3 Methodology

This study employs a comparative analysis of three different word embedding techniques for Turkish sentiment analysis. The methodology begins with preprocessing of the Winvoker Turkish sentiment analysis dataset, cleaning and preparing the text data for vectorization. Each word embedding technique generates a distinct vector representation of the words in the dataset. These vector representations are then used as input features for four different machine learning classification models: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). The performance of each model, combined with each embedding tech-

nique, is evaluated using accuracy, precision, recall, and F1-score metrics. This comparative analysis allows for a detailed assessment of the strengths and weaknesses of each word embedding method in the context of Turkish sentiment analysis and determines the most effective combination for achieving optimal classification performance.

### 3.1 Data Preprocessing

The preprocessing pipeline implemented in this study included a series of carefully designed transformations to prepare the textual data for subsequent analysis. The raw text data went through several basic procedures: Firstly, a filtering process was applied to retain only the characters of the Turkish alphabet (both uppercase and lowercase letters and additional characters specific to the Turkish orthography) and whitespace characters. All other characters, such as punctuation marks, numbers, and special symbols, were removed to reduce dimensionality and noise in the data. This decision simplifies the feature space while acknowledging the possibility of losing semantically relevant information conveyed through punctuation marks or other non-alphabetic characters.

The entire text was then converted to lowercase to standardize the word representation and eliminate case sensitivity as a differentiating factor in the analysis. This step ensured consistency and prevented the algorithm from treating changes in case usage as separate word units. Finally, leading and trailing spaces were removed from each text string to create a clean and consistent format for subsequent processing.

After text cleaning, categorical sentiment labels were converted to numerical representations using a label encoding scheme. This transformation mapped the original sentiment categories (e.g., positive, negative, neutral) to numerical values and ensured compliance with the input requirements of the machine learning models used in this study. The specific mapping used was carefully recorded and reported to ensure full transparency and ease of interpretation of the results. This coding step ensured that the classification models received numerical input for the target variable, a necessary condition for many machine learning algorithms. Morphological preprocessing steps such as stemming and lemmatization were not applied in this study, as the focus was on evaluating the embedding methods under a standardized preprocessing pipeline. The impact of such techniques on model performance remains an avenue for future research.

### 3.2 Text Vectorization

In this study, TF-IDF, Word2Vec and FastText word embedding models are compared. Based on this comparison, it is determined which embedding model is better for the dataset used.

**TF-IDF.** TF-IDF, a fundamental technique in information retrieval and text mining, represents words as vectors based on their frequency within a document and their inverse frequency across the entire corpus [3]. While simple and computationally

inexpensive, TF-IDF suffers from a significant limitation: it ignores word order and contextual information, leading to a potentially impoverished representation of sentiment. This can be particularly problematic when dealing with complex sentences, where the sentiment expressed is highly dependent on word order and the relationships between words [4].

TF-IDF takes into account both the frequency of a word in a document and the prevalence of that word in the corpus. This allows it to identify terms that are more important to the document than they would be if considered alone [22]. By identifying the most important terms, TF-IDF is a useful tool for analyzing and understanding text. This model helps to identify and prioritize the most relevant words in a text [23]. It combines two key metrics: term frequency (TF) and inverse document frequency (IDF). TF measures the frequency of a specific term within a document, normalized by the total number of terms in that document:

$$TF = \frac{Number\ of\ times\ the\ term\ appears\ in\ the\ document}{Total\ number\ of\ terms\ in\ document} \quad (1)$$

A higher TF value indicates a greater importance of the term within the specific document. IDF, conversely, measures the inverse frequency of a term across the entire corpus. It's calculated as:

$$IDF(t) = log(\frac{N}{df(t)}) \quad (2)$$

where $N$ is the total number of documents in the corpus, and $df(t)$ is the number of documents containing term $t$. A higher IDF value signifies that the term is rare across the corpus, indicating greater discriminative power. The logarithm is used to dampen the effect of very high document counts. TF-IDF combines these two metrics to generate a weighted score for each term in each document:

$$TF\text{-}IDF = TF * IDF \quad (3)$$

This score measures the significance of a term in a document based on both its frequency within that document and its relative rarity across the overall corpus. Terms that appear frequently within a particular document but are rare across the corpus will receive a high TF-IDF score, reflecting their importance in characterizing the content of that document [24]. Conversely, terms that are frequent across the entire corpus will have a lower IDF score, reducing their overall TF-IDF weight. This weighting scheme allows TF-IDF to effectively highlight terms that are both frequent within a specific document and distinctive across the broader collection, making it a powerful technique for feature extraction in text analysis tasks, such as sentiment classification [25].

**Word2Vec.** Word2Vec addresses this limitation by generating dense vector representations (embeddings) of words that capture semantic relationships between words based on their co-occurrence in a large corpus of text [5]. Word2Vec models, such as

Continuous Bag-of-Words (CBOW) and Skip-gram, learn contextual relationships, leading to more nuanced representations than TF-IDF [26]. However, Word2Vec embeddings are still context-independent, meaning that the same word has the same vector representation regardless of its usage within a sentence [7].

Word2Vec represents a family of model architectures designed to generate distributed word embeddings that map words to dense vectors of real numbers within a continuous vector space [27]. The underlying principle is that semantically similar words will exhibit proximity within this space. This is achieved through the training of shallow, two-layer neural networks on extensive textual corpora. Two primary architectures are commonly employed: CBOW and Skip-gram. CBOW predicts a target word based on its surrounding context words, while Skip-gram performs the inverse operation, predicting context words given a target word [28]. Both approaches are based on the distributional hypothesis, which states that words with similar contexts tend to have related meanings.

While a singular, concise formula cannot fully encapsulate the complexities of Word2Vec, the Skip-gram architecture can be elucidated conceptually. The objective function of Skip-gram aims to maximize the probability of observing context words given a central word.

*w(i): The central word within a given window of context*
*c(i): A context word surrounding w(i)*
*v(w(i)): The input vector representation of w(i)*
*u(c(i)): The output vector representation of c(i)* (4)

The Skip-gram model seeks to maximize the following likelihood function across the entire corpus:

$$\textit{Maximize: } \prod \Big( \big(w(i), c(i)\big) \in Corpus \Big) P(c(i)|w(i)) \tag{5}$$

The conditional probability P(c(i) | w(i)) is often modeled using a SoftMax function:

$$P(c(i)|w(i)) = \frac{exp\ (u(c(i)) \cdot v(w(i)))}{\sum (j \in Vocabulary) exp\ (u(j) \cdot v(w(i)))} \tag{6}$$

where '·' denotes the dot product. Model training involves iterative adjustment of the input and output vectors (v and u) to optimize this likelihood function, typically employing stochastic gradient descent or similar optimization algorithms. The resulting word embeddings are commonly represented by the learned input vectors (v), encapsulating the semantic relationships derived during the training process. It is important to note that this description represents a simplified representation; practical implementations often incorporate efficiency-enhancing techniques such as negative sampling or hierarchical SoftMax to address the computational challenges associated with the SoftMax normalization over the entire vocabulary. Nevertheless, this formulation provides a fundamental understanding of the mathematical underpinnings of the Skip-gram model within the Word2Vec framework.

**FastText.** FastText, building upon the strengths of Word2Vec, further refines the representation of words by incorporating subword information [29]. Unlike Word2Vec, which treats each word as an atomic unit, FastText considers the internal structure of words by generating embeddings not only for the entire word but also for its constituent n-grams (subsequences of characters). This innovative approach proves particularly advantageous when dealing with morphologically rich languages like Turkish, where a single word can encompass multiple morphemes carrying distinct semantic contributions [8]. The inclusion of n-gram embeddings allows FastText to effectively represent rare words, out-of-vocabulary terms (OOV), and morphologically complex words that might be inadequately captured by Word2Vec [30]. This enhanced representation provides a more robust and nuanced understanding of word meaning, leading to improved performance in downstream tasks such as sentiment analysis, especially when faced with the challenges presented by the complexities of Turkish grammar and vocabulary. By capturing subword information, FastText effectively bridges the gap between the limited context awareness of TF-IDF and the context-independent nature of standard Word2Vec, providing a powerful approach for creating rich and informative word embeddings [9].

Unlike Word2Vec, which assigns a single vector per word, FastText generates embeddings for both the entire word and its constituent n-grams [29, 31]. This allows FastText to handle rare words and OOV terms more effectively, a particularly valuable feature for morphologically rich languages.

While, like Word2Vec, a single formula doesn't fully capture FastText's complexity, we can outline a general framework for its objective function. Similar to Word2Vec, the goal is to learn vector representations that maximize the likelihood of observing context words given a target word.

$$\begin{gathered} V: \text{The vocabulary of words} \\ w(i): \text{The central word within a given window of context} \\ c(i): \text{A context word surrounding } w(i) \\ G(w(i)): \text{The set of n grams that constitute word } w(i). \\ \text{This includes the word itself as a special n-gram.} \\ v(g) \in R(d): \text{The d dimensional vector embedding for n gram } g. \\ \Theta: \text{The set of all model parameters} \end{gathered} \tag{7}$$

The objective function for FastText aims to maximize the log-likelihood of the observed word-context pairs in the training corpus:

$$\text{Maximize: } L(\Theta): \sum((w(i), C(i)) \in \text{Corpus}) \log P(C(i)|w(i); \Theta) \tag{8}$$

Where:

- C(i): Represents the context words associated with w(i)
- P(C(i) | w(i); Θ): The probability of observing the context words C(i) given word w(i), parameterized by Θ. This probability is calculated considering the embeddings of n grams composing w(i).

The crucial difference from Word2Vec lies in the calculation of P(C(i) | w(i); θ). Instead of relying solely on the embedding of w(i), FastText uses a combination of the embeddings of its constituent n-grams:

$$\textit{Representation of } w(i) = \sum(g \in G(w(i)))v(g) \tag{9}$$

This aggregated representation is then used in the probability calculation, typically using a SoftMax function (or its approximations, such as negative sampling) similar to Word2Vec. The model parameters θ are learned through optimization techniques like stochastic gradient descent to maximize the log-likelihood. The learned n-gram embeddings v(g) are then used to represent words and subwords, offering a more nuanced and robust representation compared to Word2Vec, particularly beneficial for morphologically rich languages and handling OOV words. The specific implementation details, such as the choice of n-gram lengths and the optimization algorithm, can affect the performance of the model.

### 3.3 Machine Learning Classifiers

**Logistic Regression (LR).** LR [32] is a statistical method used for binary and multi-class classification tasks. It models the relationship between one or more independent variables and a categorical dependent variable using a logistic function. The decision boundary is determined by a linear combination of the input features and their corresponding weights, learned during training. LR is computationally efficient, interpretable, and effective when the data exhibits a linear separable pattern. However, its performance can degrade in cases of multicollinearity or non-linearity without appropriate preprocessing or feature engineering.

**Decision Tree (DT).** A DT [33] is a non-parametric, tree-structured algorithm used for both classification and regression tasks. It divides the feature space into regions based on recursive binary splits guided by a criterion such as Gini impurity or information gain. Each internal node represents a decision based on a single feature, while the leaf nodes correspond to the predicted results. DTs are intuitive, easy to visualize, and can handle non-linear relationships effectively. However, they are prone to overfitting, especially when the tree depth is not restricted, and their performance can be sensitive to small variations in the data.

**Random Forest (RF).** RF [34] is an ensemble learning method that builds multiple decision trees during training and aggregates their predictions to improve classification or regression performance. Each tree is constructed using a random subset of the training data and a random selection of features, which introduces diversity and reduces overfitting. The final prediction is typically determined through majority voting for classification or averaging for regression tasks. RF models are robust to noise, can handle high-dimensional data, and are relatively immune to overfitting. However, they can be computationally intensive for large datasets and lack interpretability compared to simpler models.

**Support Vector Machine (SVM).** SVM [35] is a supervised learning algorithm that aims to find the optimal hyperplane to separate data points belonging to different classes with the maximum margin. It uses kernel functions, such as linear, polynomial, or radial basis functions, to handle non-linear separability by transforming the input space into a higher-dimensional feature space. SVMs are particularly effective for datasets with clear margins of separation and are less susceptible to overfitting in high-dimensional spaces. However, they require careful parameter tuning and can be computationally expensive for large datasets.

## 4 Experimental Results

The optimum parameters of the classifiers used in the experiments were carefully determined to maximize the performance of each model and are presented in Table 1. In this parameter optimization process, hyperparameters for each classifier were examined with a comprehensive screening method and the combinations that provided the best results were selected. This process is a critical step to increase model accuracy and ensure the reliability of the experimental results, while also ensuring that the comparison between different classifiers is fair. The parameter values listed in Table 1 represent the optimized results according to the structures of the relevant algorithms and the dataset.

**Table 1.** Optimal Parameters for Classifiers

| Model | Parameter | Value |
|---|---|---|
| TF-IDF | max_features | 10000 |
| | ngram_range | (1, 2) |
| | stop_words | None |
| Word2Vec | vector_size | 300 |
| | window | 5 |
| | min_count | 1 |
| | workers | 4 |
| FastText | vector_size | 100 |
| | window | 5 |
| | min_count | 5 |
| | workers | 3 |
| | word_ngrams | 1 |
| LR | max_iter | 1000 |
| RF | n_estimators | 10 |
| SVM | max_iter | 1000 |

The analyses we conducted in this study were performed using the Winvoker dataset. Winvoker's training dataset consists of 489,644 instances of 262,166 positive, 56,561 negative and 170,917 neutral classes, compiled from different sources such as humir,

e-commerce site comments and Wikipedia. There are a total of 48965 examples in the test data set. The class distribution in the test data set is as shown in Fig 1. Most sentiment models have only two labels; positive and negative. However, the user input can be entirely notr sentences. Positive and negative sentences were taken from comments, while notr sentences were extracted from the Turkish wiki transcript. Table 2 shows an example of each class.

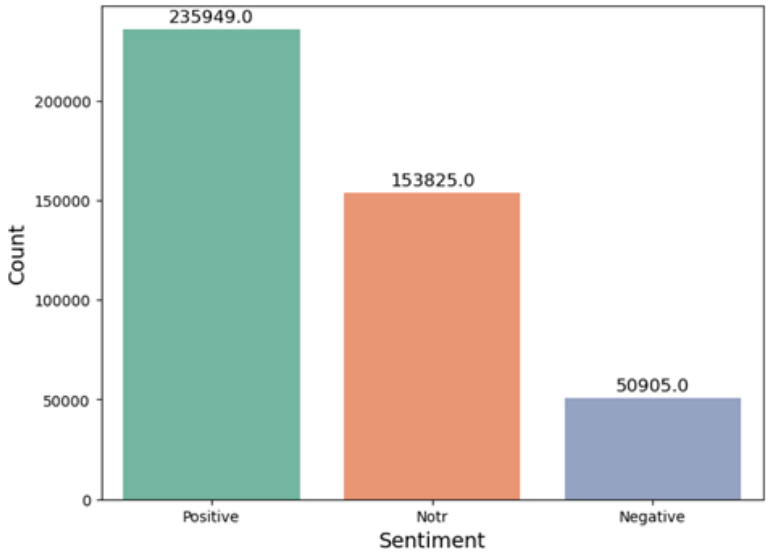



**Fig. 1.** Class distribution in test dataset.

**Table 2.** Data examples of classes from the dataset.

| Text | Label | Dataset |
|---|---|---|
| hızlı kargo, temiz alışveriş. teşekkür ederim. | Positive | urun_yorumlari |
| Çünkü aranan tapınak bu bölgededir. | Notr | wiki |
| özellikle ışığı çok yetersiz ve 4 tane değil 2 tane geliyor bana siyah geldi ancak ışığı dediğim gibi yetersiz | Negative | urun_yorumlari |

The performance of the proposed models was evaluated using a set of standard classification metrics, ensuring a comprehensive assessment. Accuracy (10) was employed as a primary metric, representing the proportion of correctly classified instances to the total number of instances, providing an overall measure of the model's predictive capability. Precision (11), defined as the proportion of true positive predictions out of all positive predictions, was used to evaluate the model's ability to minimize false positive rates, which is critical in scenarios where false alarms carry significant consequences.

In addition to precision, recall (12) was utilized to evaluate the model's efficacy in identifying actual positive cases, a particularly salient aspect in tasks were failing to detect positives can be costly. The F1-score (13), representing a harmonic mean of precision and recall, was incorporated to balance the trade-off between these two metrics, a strategy that is especially valuable for imbalanced datasets. The confusion matrix, a tool that provides a detailed breakdown of the model's predictions, was also employed to illuminate the model's strengths and weaknesses. This comprehensive analysis, facilitated by the incorporation of various metrics, ensured a robust and multi-faceted evaluation of model performance.

$$Accuracy = (TP+TN)/(TP+TN+FN+FP) \quad (10)$$

$$Precision = TP/(TP+FP) \quad (11)$$

$$Recall = TP/(TP+FN) \quad (12)$$

$$F1{-}Score = 2*((Precision*Recall)/(Precision+Recall)) \quad (13)$$

### 4.1 Results

To compare the performance of the three word embedding techniques (TF-IDF, Word2Vec, and FastText) in the context of Turkish sentiment analysis, we trained four different machine learning models: LR, DT, RF, and SVM. Each embedding technique was used with each model, resulting in twelve separate model configurations. The following metrics were calculated for each configuration: Accuracy, precision, recall, f1-score, and time. The results are summarized in Table 3.

**Table 3.** Performance of Sentiment Analysis Models Across Embedding Techniques

| Embedding | Model | Accuracy | Precision | Recall | F1-Score | Time |
|---|---|---|---|---|---|---|
| TF-IDF | LR | 0.633 | 0.60 | 0.63 | 0.61 | 59.74 |
| TF-IDF | DT | 0.585 | 0.58 | 0.59 | 0.58 | 410.19 |
| TF-IDF | RF | 0.589 | 0.58 | 0.59 | 0.58 | 199.50 |
| TF-IDF | SVM | 0.376 | 0.54 | 0.38 | 0.28 | 73.60 |
| **Word2Vec** | **LR** | **0.911** | **0.91** | **0.91** | **0.91** | **390.08** |
| Word2Vec | DT | 0.845 | 0.85 | 0.85 | 0.85 | 412.84 |
| Word2Vec | RF | 0.895 | 0.89 | 0.90 | 0.89 | 97.09 |
| Word2Vec | SVM | 0.331 | 0.31 | 0.33 | 0.32 | 443.42 |
| FastText | LR | 0.767 | 0.75 | 0.77 | 0.75 | 225.36 |
| FastText | DT | 0.683 | 0.69 | 0.68 | 0.68 | 265.06 |
| FastText | RF | 0.760 | 0.74 | 0.76 | 0.75 | 89.59 |
| FastText | SVM | 0.554 | 0.59 | 0.55 | 0.48 | 440.97 |

The results show a clear performance advantage for Word2Vec in at least three of the four machine learning models. In particular, LR yielded the highest accuracy scores across all three embedding methods. With Word2Vec, LR reached an accuracy of 0.911, significantly outperforming other models using Word2Vec.

The results clearly demonstrate the superiority of Word2Vec and, to a lesser extent, FastText over TF-IDF. TF-IDF, a simple bag-of-words approach, relies solely on word frequencies and ignores word order and semantic relationships. This limitation is reflected in its significantly lower accuracy scores across all models compared to the other two embedding methods. The average accuracy for TF-IDF is approximately 0.54, while Word2Vec achieves an average accuracy of roughly 0.79 and FastText an average of 0.70. Word2Vec and FastText, on the other hand, generate dense vector representations of words that capture semantic similarities. This leads to better performance in downstream tasks like text classification. Word2Vec consistently outperforms FastText, suggesting that its approach to generating word embeddings is more effective for this specific dataset and task. However, the training time for Word2Vec is noticeably longer, particularly for LR and DT models. FastText provides a good balance between accuracy and speed, making it a potentially attractive option when computational resources are limited.

Comparison of the machine learning models reveals distinct performance characteristics. LR achieves remarkable accuracy (0.911) when paired with Word2Vec embeddings, significantly outperforming other models with the same embedding. This suggests that Word2Vec effectively captures features highly relevant to LR's linear decision boundary. However, its performance drops substantially with TF-IDF. DTs show relatively consistent performance across embeddings, but generally lower accuracy than LR and RF, especially with TF-IDF, and high training times across the board. RF consistently outperforms DTs, showcasing the benefits of bagging in enhancing accuracy and robustness.

Similar to LR, it benefits significantly from Word2Vec embeddings, achieving high accuracy with relatively faster training times compared to DT, second only to LR with Word2Vec. In contrast, SVM consistently underperforms across all embeddings, displaying consistently low accuracy regardless of the embedding method. This suggests that SVM struggles to effectively capture the relevant features for this specific classification problem, with its chosen parameters.

The anomalously low SVM performance, particularly with Word2Vec embeddings (accuracy: 0.331), can be attributed to several factors. First, SVM is known to be sensitive to high-dimensional dense vectors; Word2Vec produces 300-dimensional dense embeddings, which may cause the algorithm to struggle with finding an optimal hyperplane. Second, the max_iter=1000 setting may have been insufficient for convergence given the large dataset size (~490K training instances), meaning the optimizer did not reach a stable solution. Third, SVMs are computationally expensive for large-scale datasets, and the long training times observed (443 seconds with Word2Vec) further support the hypothesis that the model did not fully converge. Future work could explore kernel tuning and increased iteration limits to potentially improve SVM performance on this task.

The unexpected superiority of Word2Vec over FastText in this study can be attributed to the nature of the sentiment analysis task. Sentiment classification relies primarily on whole-word semantic representations, where the overall meaning of a word carries more discriminative power than its constituent subword components. FastText's subword information, while beneficial for morphologically rich languages

in tasks such as named entity recognition or text classification involving rare words, may introduce noise in sentiment-oriented tasks where the emotional polarity is conveyed at the word level rather than the subword level. Consequently, Word2Vec's cleaner word-level representations may align more naturally with the linear decision boundaries learned by Logistic Regression, explaining its superior performance on this dataset.

The high training times for SVM with Word2Vec and FastText further highlight this model's inefficiency for this task. To further investigate model behavior, Fig 2 presents the confusion matrices for LR, using each of the three embedding methods.

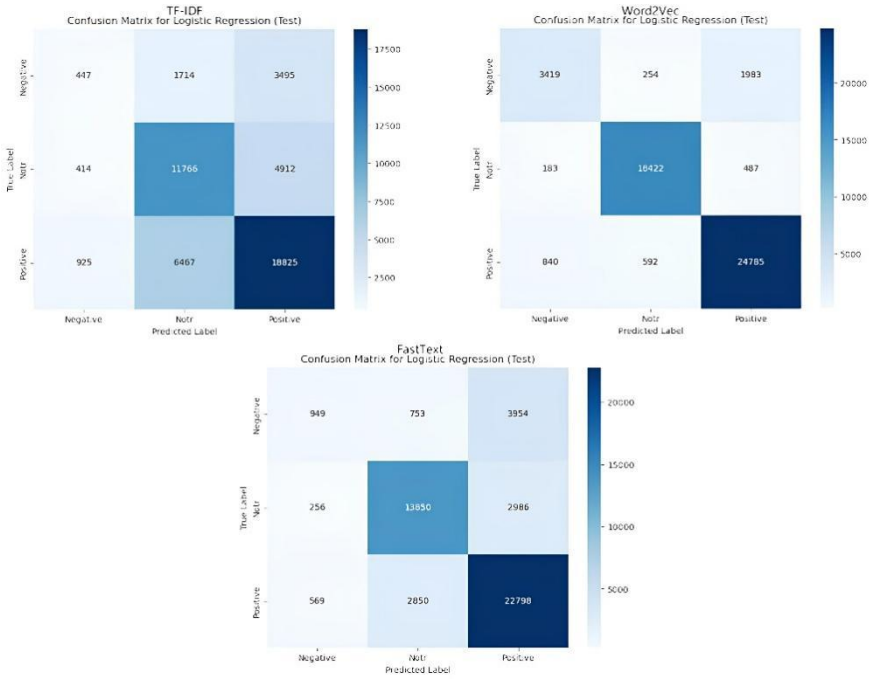



**Fig. 2.** Confusion matrices for LR with TF-IDF, Word2Vec, and FastText embeddings.

The choice of both embedding method and machine learning model significantly impacts the performance of the text classification system. Word2Vec demonstrates superior performance as an embedding method, producing significantly higher accuracy scores across most models. However, this comes at the cost of increased training time. FastText provides a good compromise between accuracy and speed. Among the machine learning models, LR and RF achieve the highest accuracy, particularly when combined with Word2Vec embeddings. However, SVM consistently underperforms, suggesting that it may not be the most appropriate model for this particular task and dataset. Further investigation could explore hyperparameter tuning for each model to potentially improve performance, particularly for SVM. Furthermore, understanding the characteristics of the dataset and choosing the appropriate evaluation metrics is

critical for a robust and informed model selection. Fig 3 provides a visual comparison of the accuracy achieved by each model for the different embedding techniques.

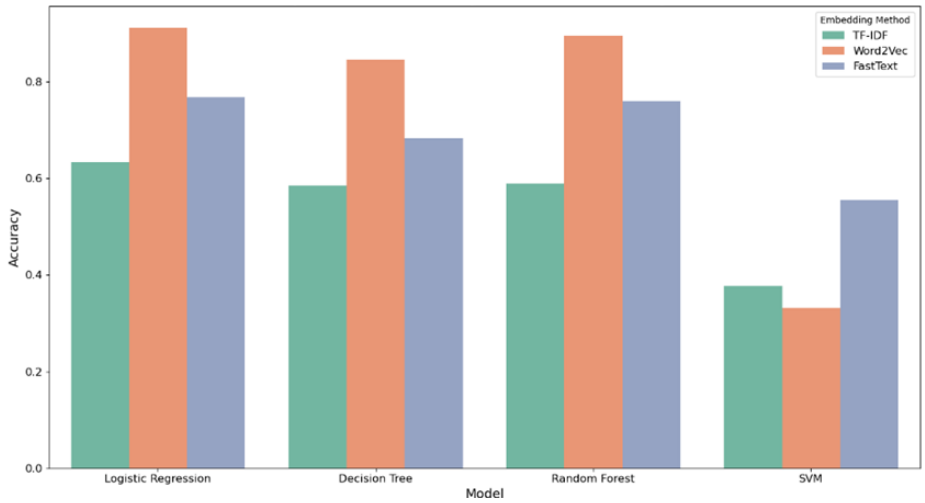



**Fig. 3.** Accuracy comparison of models across different embedding techniques.

## 5 Conclusion

This study performed a comprehensive comparative analysis of three prominent word embedding techniques (TF-IDF, Word2Vec and FastText) for Turkish sentiment analysis using four different machine learning models. Our results show that Word2Vec consistently outperforms TF-IDF and is comparable to FastText across multiple classifiers. The superior performance of Word2Vec and FastText, especially compared to the simpler TF-IDF approach, highlights the importance of considering contextual information and subword morphology to achieve high accuracy in sentiment classification for morphologically rich languages. While the context-independent nature of TF-IDF limited its effectiveness, the capacity of Word2Vec and FastText to capture semantic relations and subword structures significantly improved the classification accuracy. When comparing Word2Vec and FastText, Word2Vec performed better. This outcome suggests that for sentiment analysis tasks, word-level semantic representations may be more effective than subword-based approaches, as sentiment polarity is predominantly carried at the word level rather than the subword level. Future research could investigate the effectiveness of incorporating advanced techniques such as attention mechanisms or transducer-based architectures to further improve the performance of sentiment analysis models in Turkish. Furthermore, further research on larger and more diverse datasets may help to verify the generalizability of our findings and improve the selection of the most appropriate embedding techniques for different applications. The limitations of our study, mainly related to the specific dataset and the selected classifiers, should be taken into account when interpreting the results and guiding future research directions.

## References


1. Atf, Z., Taherikia, F., & Heidarzadeh Hanzaee, K. (2023). Modeling Brand Engagement in Social Media (Based on Sentiment Analysis and Customer Data). International Journal of Innovation Management and Organizational Behavior (IJIMOB), 3(2), 208-218.
2. Farias, D. H., & Rosso, P. (2017). Irony, sarcasm, and sentiment analysis. In Sentiment analysis in social networks (pp. 113-128). Morgan Kaufmann.
3. Mohd Sofi, S., & Selamat, A. (2023). Aspect Based Sentiment Analysis: Feature Extraction using Latent Dirichlet Allocation (LDA) and Term Frequency - Inverse Document Frequency (TF-IDF) in Machine Learning (ML). Malaysian Journal of Information and Communication Technology (MyJICT), 8(2), 158-168. https://doi.org/10.53840/myjict8-2-102
4. Wu, H., & Yuan, N. (2018, May). An Improved TF-IDF algorithm based on word frequency distribution information and category distribution information. In Proceedings of the 3rd International Conference on Intelligent Information Processing (pp. 211-215)
5. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
6. Sivakumar, S., Videla, L. S., Kumar, T. R., Nagaraj, J., Itnal, S., & Haritha, D. (2020, September). Review on word2vec word embedding neural net. In 2020 international conference on smart electronics and communication (ICOSEC) (pp. 282-290). IEEE.
7. Sharma, A. (2023). Context-aware Sentiment Analysis on Refined Word Embeddings Word2Vec Model. Authorea Preprints.
8. Vasić, D., & Brajković, E. (2018, September). Development and evaluation of word embeddings for morphologically rich languages. In 2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-5). IEEE.
9. Savelli, C., & Giobergia, F. (2024, September). Enhancing Cross-Lingual Word Embeddings: Aligned Subword Vectors for Out-of-Vocabulary Terms in fastText. In 2024 IEEE 18th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-6). IEEE.
10. "winvoker/turkish-sentiment-analysis-dataset · Datasets at Hugging Face", 09 Jan 2025, https://huggingface.co/datasets/winvoker/turkish-sentiment-analysis-dataset
11. Demircan, M., Seller, A., Abut, F., & Akay, M. F. (2021). Developing Turkish sentiment analysis models using machine learning and e-commerce data. International Journal of Cognitive Computing in Engineering, 2, 202-207.
12. Aydın, C. R., & Güngör, T. (2021). Sentiment analysis in Turkish: Supervised, semi-supervised, and unsupervised techniques. Natural Language Engineering, 27(4), 455-483.
13. Kemaloğlu, N., Küçüksille, E., & Özgünsür, M. (2021). Turkish sentiment analysis on social media. Sakarya University Journal of Science, 25(3), 629-638.
14. Özdemir, M., & Ortakcı, Y. (2024). Text Classification with Frequency-Based Text Vectorisation Methods for Enhancing Call Centre Efficiency. Current Trends in Computing, 1(2), 122-138.
15. Acikalin, U. U., Bardak, B., & Kutlu, M. (2020, October). Turkish sentiment analysis using Bert. In 2020 28th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
16. Guven, Z. A. (2022). The comparison of language models with a novel text filtering approach for turkish sentiment analysis. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(2), 1-16.

17. Ba Alawi, A., & Bozkurt, F. (2024). Performance Analysis of embedding methods for deep learning-based Turkish sentiment analysis models. Arabian Journal for Science and Engineering, 1-23.
18. Alshammari, Q., & Akyüz, S. (2024). Mental Health on Twitter in Turkey: Sentiment Analysis with Transformers. In Decision Making in Healthcare Systems (pp. 391-402). Cham: Springer International Publishing.
19. Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. Heliyon, 10(1).
20. Taşar, D. E., Özcan, C., & Koruyan, K. (2022). Autotrain Yaklaşimi İle Duygu Analizi. Proceeding Book, 30.
21. Özdemir, M., & Ortakcı, Y. (2024). An Analysis of Intelligent Turkish Text Classification Models for Routing Calls in Call Centers: A Case Study on the Republic of Türkiye Ministry of Trade Call Center. Sakarya University Journal of Computer and Information Sciences, 7(1), 46-60.
22. Bui, Q. A., Visani, M., Prum, S., & Ogier, J. M. (2011, September). Writer identification using TF-IDF for cursive handwritten word recognition. In 2011 International Conference on Document Analysis and Recognition (pp. 844-848). IEEE.
23. Naeem, M. Z., Rustam, F., Mehmood, A., Ashraf, I., & Choi, G. S. (2022). Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. PeerJ Computer Science, 8, e914.
24. Alfarizi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional text classification using tf-idf (term frequency-inverse document frequency) and lstm (long short-term memory). JUITA: Jurnal Informatika, 10(2), 225-232.
25. Alshuraiqi, H. (2020). Improved term frequency inverse document frequency (TF-IDF) method for Arabic text classification. International Journal of Advanced Trends in Computer Science and Engineering, 9(1), 6939–6946.
26. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems, 26.
27. Jatnika, D., Bijaksana, M. A., & Suryani, A. A. (2019). Word2vec model analysis for semantic similarities in english words. Procedia Computer Science, 157, 160-167.
28. Li, P. (2023). The study for word prediction based on Skip-gram and CBOW model. Theoretical and Natural Science, 18(1), 210–215.
29. Kang, H., & Yang, J. (2020). Performance comparison of Word2vec and fastText embedding models. Journal of Digital Contents Society, 21(7), 1335-1343.
30. Liu, K. (2021). Incorporate Out-of-Vocabulary Words for Psycholinguistic Analysis using Social Media Texts-An OOV-Aware Data Curation Process and a Hybrid Approach (Doctoral dissertation, The Claremont Graduate University).
31. Nasution, N., Nababan, E., & Mawengkang, H. (2024). Comparing LSTM algorithm with word embedding: FastText and Word2Vec in Bahasa Batak-English translation. Proceedings of the International Conference on Information and Communication Technology (ICoICT), 306–313.
32. Das, A. (2024). Logistic regression. In Encyclopedia of Quality of Life and Well-Being Research (pp. 3985-3986). Cham: Springer International Publishing.
33. De Ville, B. (2013). Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics, 5(6), 448-455.
34. Iranzad, R., & Liu, X. (2024). A review of random forest-based feature selection methods for data science education and applications. International Journal of Data Science and Analytics, 1-15.

35. Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. International Journal of Information Technology, 13(5), 1-11.

# Industry and Technical Tools section

# Smart Buildings, Smart Villages, and Smart Cities: Applications of Digital Technologies in Architecture

Gafarbayli Konul, Aslanov Bayram
Azerbaijan University of Architecture and Construction, Baki, Azerbaijan,
qafarbeyli_konul@mail.ru

**Abstract**

The global transition toward digital transformation has revolutionized architecture and construction, paving the way for intelligent systems that improve sustainability, efficiency, and livability. Smart buildings, smart villages, and smart cities represent an interconnected framework where architecture integrates information technologies such as the Internet of Things (IoT), Building Information Modeling (BIM), artificial intelligence (AI), and data analytics. This paper explores the application of these technologies in the field of architecture, focusing on energy management, sustainable urban planning, and digital infrastructure. The study highlights examples from both global and Azerbaijani contexts, including the "Aghaly Smart Village" project in Zangilan, which stands as a leading model for post-conflict reconstruction and rural innovation.

***Keywords:*** Smart building, smart city, smart village, architecture, digital technologies, BIM, IoT, Azerbaijan

## 1. Introduction

In the twenty-first century, information technologies have become a key driver of change in architecture and construction. The rise of intelligent systems has transformed how buildings are designed, constructed, and operated, introducing a new paradigm known as smart architecture. The concept of "smartness" in the built environment refers to the integration of digital systems that collect, analyze, and act upon real-time data to optimize energy use, enhance comfort, and ensure environmental sustainability.

According to the United Nations (2024), by 2050 more than 68% of the global population will live in urban areas. This demographic trend has intensified the need for innovative urban planning approaches that ensure sustainable living conditions while minimizing environmental impact. As a result, smart city initiatives have become a global priority, combining advanced technologies, data management systems, and participatory governance to improve urban life.

In architecture, the application of information technologies such as Building Information Modeling (BIM), Geographic Information Systems (GIS), and IoT-based automation enables a new level of design precision, performance monitoring, and resource management. Smart buildings utilize automated systems to manage heating, ventilation, lighting, and security, while smart villages and cities apply similar principles to larger-scale planning and infrastructure management. These technologies

help architects and planners create adaptive environments that respond to users' needs and external conditions in real time.

## 2. Background and Conceptual Framework

The term smart in the context of architecture goes beyond mere technological integration—it represents a shift toward data-driven and sustainable design. Smart buildings and urban environments are equipped with sensors, control systems, and communication networks that enable real-time monitoring and decision-making. These technologies contribute to improved energy performance, reduced emissions, and enhanced quality of life for residents.

### 2.1 Smart Buildings

Smart buildings combine architectural design with digital intelligence. Through the use of IoT sensors, BIM models, and AI-driven analytics, these structures continuously monitor internal and external environmental parameters. For instance, temperature and occupancy sensors allow automated adjustment of lighting and air conditioning, ensuring both comfort and energy efficiency. According to Statista (2024), smart building technologies can reduce energy consumption by up to 30–35%, significantly lowering operational costs.

In architectural practice, smart building systems enable adaptive façades, real-time performance analysis, and integration of renewable energy sources such as solar panels and wind turbines.

### 2.2 Smart Villages

The smart village concept has gained momentum as rural areas seek to bridge the digital divide. A smart village integrates communication technologies, renewable energy systems, and digital management platforms to improve local governance and sustainability. In Azerbaijan, the "Aghaly Smart Village" in Zangilan district represents the first pilot project of its kind in the South Caucasus. The project includes smart homes equipped with solar energy, e-government services, waste management systems, and digital agricultural monitoring. It stands as an example of how post-conflict regions can leverage innovation to ensure sustainable reconstruction and rural revitalization.

### 2.3 Smart Cities

Smart cities extend the principles of intelligent systems to the urban scale. They use digital networks to manage transportation, water supply, waste disposal, and environmental quality. For example, Singapore, Dubai, and Amsterdam have implemented advanced smart city strategies that integrate data analytics, renewable energy, and citizen participation.

According to the IMD Smart City Index (2024), cities that adopt data-based governance and digital infrastructure report higher sustainability indicators and citizen satisfaction levels.

## 3. Technological Tools and Innovations in Architecture

The integration of digital technologies into architectural practice has reshaped the way designers, engineers, and urban planners approach the built environment. Among the most significant innovations are Building Information Modeling (BIM), the Internet of Things (IoT), Artificial Intelligence (AI), and Geographic Information Systems (GIS). Together, these systems enable architects to create data-informed designs that enhance performance, sustainability, and user experience.

### 3.1 Building Information Modeling (BIM)

BIM represents a fundamental shift in how construction projects are conceived and executed. It allows architects and engineers to work within a single, shared 3D digital model that integrates architectural, structural, and mechanical systems.

According to Autodesk (2024), BIM adoption in the global construction industry has surpassed 70%, and in countries such as the United Kingdom and Singapore, it is mandatory for public projects. BIM models enhance collaboration, reduce errors, and minimize construction waste by enabling precise simulation and planning.

In the context of smart buildings, BIM serves as the foundation for digital twins — virtual replicas that collect and process data from real-time sensors embedded within the structure. This integration allows for predictive maintenance, energy optimization, and improved occupant comfort.

### 3.2 Internet of Things (IoT) in Architecture

IoT devices are transforming architectural environments into responsive, intelligent ecosystems. By connecting sensors, cameras, and controllers through the internet, buildings and urban systems can “communicate” with one another.

For example, smart lighting systems adjust brightness according to daylight availability and occupancy levels, while HVAC systems automatically regulate temperature and air quality. Studies show that IoT integration can reduce energy consumption by 20–40% in commercial and residential buildings (Statista, 2024).

In urban settings, IoT networks manage traffic flow, monitor air pollution, and ensure efficient waste collection. These applications not only improve living standards but also contribute to climate resilience and sustainable city development.

### 3.3 Artificial Intelligence (AI) and Data Analytics

AI plays a key role in analyzing the massive volumes of data generated by smart systems. Machine learning algorithms can predict building energy use, simulate

architectural forms, and identify optimal design solutions. AI-driven design tools such as Generative Design allow architects to explore thousands of structural variations based on sustainability and material efficiency criteria.

In cities like Dubai, AI platforms are used to manage infrastructure maintenance and optimize transportation routes, leading to significant reductions in carbon emissions and traffic congestion.

### 3.4 Geographic Information Systems (GIS)

GIS technology supports spatial analysis and planning in both rural and urban contexts. It provides a visual and analytical framework for mapping resources, population density, and land use. In architecture, GIS is particularly valuable for site selection, environmental assessment, and disaster management planning.

For Azerbaijan, GIS integration into national planning policies could enhance reconstruction projects in liberated territories, including Aghdam and Zangilan, by supporting sustainable land-use strategies and spatial data management.

## 4. Case Studies

### 4.1 The Aghaly Smart Village, Azerbaijan

The Aghaly Smart Village, located in the Zangilan district of Azerbaijan, is one of the most significant examples of smart rural reconstruction in the post-conflict region of the South Caucasus. Implemented under the “Smart Village” concept initiated by the Government of Azerbaijan, the project demonstrates how digital technologies can revitalize depopulated rural areas through sustainable and intelligent design.

Aghaly integrates renewable energy systems, energy-efficient housing, solar power generation, electronic governance services, automated waste management, and smart agriculture. Each house is equipped with modern IoT-based metering and communication systems. The project emphasizes not only technological innovation but also social reintegration, creating a model for rural development based on sustainability and community well-being.

According to the Azerbaijan Ministry of Energy (2024), Aghaly’s renewable energy infrastructure supplies over 40% of local energy demand, while digital monitoring systems have reduced water and energy waste by approximately 30% compared to conventional rural settlements.

## Conclusion

The transformation toward smart and sustainable cities represents one of the most ambitious undertakings of the 21st century — a movement where technology, environmental awareness, and cultural heritage converge to redefine how societies evolve. Across Azerbaijan, this shift is not merely about modernization, but about creating a resilient framework where innovation serves humanity and nature alike. Smart urbanism has become a strategic pathway for aligning national growth with

global sustainability goals, ensuring that digital progress goes hand in hand with ecological responsibility.

Architecture and urban planning now function as intelligent systems — capable of learning, adapting, and responding to both environmental and social challenges. The integration of AI, GIS mapping, and data-driven planning transforms cities into living organisms that monitor and optimize their own resource flows. Through green construction materials, renewable energy, and circular waste systems, the built environment evolves into a self-sustaining ecosystem. Every structure, every public space, and every network becomes a node of resilience in a larger sustainable system.

Karabakh, in this context, stands as the most visible and symbolic expression of Azerbaijan's smart development vision. The post-war reconstruction process has turned the region into a vast testing ground for innovation, sustainability, and digital design. Cities such as Shusha, Fuzuli, Jabrayil, Lachin, and Kalbajar illustrate how technology can revive not only infrastructure but also identity, culture, and community life. Through digital twins, renewable energy integration, and advanced resource management, these areas are being reborn as intelligent, green, and culturally rooted spaces. The region demonstrates that rebuilding can mean more than recovery — it can mean reimagining the future.

Yet, the true success of a smart region lies not in its technology alone, but in its people. Human-centered design, citizen engagement platforms, and inclusive governance models ensure that innovation remains accessible, equitable, and responsive to real human needs. Education, digital literacy, and sustainable awareness are the foundations upon which smart societies are built. They guarantee that the benefits of innovation extend to all — from urban centers to rural communities — creating a sense of shared progress and common responsibility.

As Azerbaijan advances on its path toward a data-driven, low-carbon, and culturally vibrant future, its experience becomes a blueprint for others. The lessons from smart reconstruction and green regional planning extend far beyond national borders. They demonstrate that progress can honor tradition, and that the cities of tomorrow can be both technologically advanced and profoundly human.

In the end, the journey toward a sustainable smart region is not a destination, but an evolving vision — one that continuously adapts, learns, and inspires. It is a vision where architecture becomes a dialogue between past and future, where technology amplifies culture rather than replaces it, and where every innovation serves a single purpose: to create harmony between people, nature, and the spaces they share.

### 4.2 Dubai Smart City Initiative

Dubai stands at the forefront of smart city development in the Middle East. Its Dubai Smart 2025 Strategy aims to transform the emirate into a fully connected digital ecosystem where governance, mobility, and construction are seamlessly integrated.

The Smart Dubai Platform collects and analyzes data from various sectors, including energy, transport, and housing. Architectural innovations such as 3D-printed construction and AI-based energy monitoring have positioned Dubai as a leader in digital urbanism.

The city's Museum of the Future and Dubai Sustainable City are architectural landmarks that reflect the integration of smart technologies into design, with real-time energy monitoring, recycled water systems, and automated climate control.

**4.3 Singapore – A Global Model of Smart Urbanism**

Singapore's Smart Nation Initiative demonstrates a comprehensive approach to urban intelligence. Through the use of IoT, AI, and big data, the city-state has achieved high efficiency in transport management, waste recycling, and urban housing.

Singapore uses BIM-based master planning across the construction sector, while its Virtual Singapore platform serves as a 3D digital twin of the entire city, allowing real-time visualization of environmental and infrastructural data.

This digital infrastructure enables planners and architects to simulate urban development scenarios and assess sustainability impacts before implementation.

**5. Statistical Overview and Analysis**

The global market for smart city and smart building technologies has grown rapidly over the last decade. According to MarketsandMarkets (2024), the smart city market is projected to reach 1.2 trillion USD by 2025, with an annual growth rate exceeding 15%. The smart building sector alone is expected to exceed 200 billion USD, driven by urbanization, sustainability targets, and energy efficiency regulations.

| Indicator | 2020 | 2024 | 2025 (Forecast) |
|---|---|---|---|
| Global Smart City Market (USD) | 480 billion | 900 billion | 1.2 trillion |
| BIM Adoption in Architecture (%) | 45% | 70% | 80% |
| Average Energy Savings from Smart Systems (%) | 15–25 | 25–35 | 30–40 |
| Share of Renewable Energy in Aghaly Village (%) | – | 40 | 50 (Goal) |

These figures underline the accelerating pace of digital transformation across architectural and urban environments. Smart technologies are not merely tools but strategic enablers of sustainability, resilience, and innovation.

For Azerbaijan, the application of smart building and city technologies in reconstruction and urban renewal offers significant potential — particularly in liberated territories and rapidly growing cities like Baku and Sumgait. The integration of BIM, IoT, and renewable energy systems can enhance the efficiency of infrastructure projects, improve public services, and ensure environmental sustainability in line with UN Sustainable Development Goals (SDG 9 and SDG 11).

3. Technological Tools and Innovations in Architecture

The integration of digital technologies into architectural practice has reshaped the way designers, engineers, and urban planners approach the built environment. Among the

most significant innovations are Building Information Modeling (BIM), the Internet of Things (IoT), Artificial Intelligence (AI), and Geographic Information Systems (GIS). Together, these systems enable architects to create data-informed designs that enhance performance, sustainability, and user experience.

**Figure 1."Smart Village" project, built on the territory of village of Aghali region. Reconstruction Zangilan Karabakh**

## 6. Smart Strategies in

### 6.1 Digital Master Planning and Smart Urban Design

The post-war reconstruction of the Karabakh region represents one of the most ambitious urban development programs in modern Azerbaijan. As the country reclaims and rebuilds its liberated territories, it is simultaneously establishing a model of digital urbanism, where technology, sustainability, and culture converge. The government's approach positions Karabakh as a "living laboratory" for innovative architectural and planning practices, guided by the principles of resilience, efficiency, and inclusivity.

The foundation of Karabakh's smart reconstruction begins with digital master planning, which integrates satellite mapping, 3D modeling, and Building Information Modeling (BIM) to reimagine cities from the ground up. These digital tools enable architects and planners to simulate various urban scenarios, analyze spatial efficiency, and optimize land use. Through Geographic Information Systems (GIS), every design decision— from the placement of residential zones to the alignment of infrastructure—is supported by real-time environmental and topographic data. This ensures that new urban areas such as Fuzuli, Jabrayil, and Shusha are developed in harmony with natural landscapes, minimizing risks of overdevelopment or environmental degradation.

### 6.2 Integration of Smart Infrastructure and Energy Systems

A crucial component of the Karabakh reconstruction process is the integration of smart infrastructure aimed at creating energy-efficient and climate-resilient environments. Urban developments incorporate IoT-based energy monitoring systems, smart grids, and renewable power sources such as solar and hydropower. Smart buildings within these areas are designed with automated lighting, temperature regulation, and water-saving technologies to reduce operational emissions. Moreover, advanced waste management systems are being implemented, where sensors track waste collection patterns to improve recycling efficiency. Such measures align with

Azerbaijan's broader commitment to the "Green Energy Zone" concept, which seeks to transform Karabakh into a model of low-carbon urbanization.

**6.3 Data-Driven Governance and Environmental Monitoring**

Smart reconstruction in Karabakh also emphasizes the use of data-driven governance to ensure transparency, sustainability, and long-term adaptability. Through the use of digital dashboards and real-time environmental monitoring platforms, local authorities can track construction progress, manage resources, and assess the ecological footprint of urban expansion. These systems collect and analyze data on air quality, energy consumption, and waste generation, allowing decision-makers to make informed adjustments to policies and design strategies. Furthermore, digital participation tools are being introduced to engage residents in urban management, ensuring that the new communities are not only technologically advanced but also socially inclusive and responsive to citizen needs.

**7. Shusha – The Smart Cultural Capital**

**7.1 Digital Heritage Preservation and Architectural Reconstruction**

The city of Shusha, often referred to as the cultural heart of Azerbaijan, is undergoing an extensive process of digital preservation and reconstruction. As part of the national "smart cultural city" framework, advanced technologies such as 3D laser scanning, photogrammetry, and AI-based architectural modeling are employed to document and restore the city's historical monuments, mosques, and residential buildings with high precision. Each restoration project begins with a detailed digital survey that captures the architectural geometry, material condition, and historical details of structures. This digital heritage database not only guides restoration work but also serves as an educational and archival resource for future generations. Through these technologies, Shusha's urban identity is being rebuilt in a way that preserves authenticity while embracing the possibilities of digital innovation.

**7.2 Integration of Smart Infrastructure and Sustainable Urban Systems**

Alongside cultural restoration, Shusha's urban renewal integrates smart and sustainable infrastructure systems that redefine the relationship between heritage and modern living. The city's energy network is designed to operate through renewable sources, including solar microgrids and energy-efficient public lighting systems that automatically adjust based on time and weather conditions. IoT-based environmental sensors continuously monitor air quality, humidity, and energy consumption, ensuring that the ecological footprint of restoration and daily activities remains minimal. Smart water distribution systems and digital waste management solutions contribute to resource efficiency and improve residents' quality of life. By combining cultural continuity with ecological awareness, Shusha stands as a model for how technology can coexist with—and even enhance—the preservation of historical environments.

**7.3 Smart Tourism, Cultural Experience, and Community Engagement**

Shusha's transformation also introduces the concept of **smart tourism**, where digital platforms enhance visitor experience while promoting cultural sustainability. Interactive applications and augmented reality (AR) tools allow tourists to explore reconstructed landmarks, such as the Yukhari Govhar Agha Mosque and Shusha Fortress, in both their historical and modern contexts. These technologies offer virtual tours, immersive storytelling, and multilingual guidance that deepen public appreciation of the city's heritage. Furthermore, digital participation platforms enable citizens to contribute feedback on urban development plans, fostering a sense of community ownership. This combination of smart tourism, public engagement, and cultural education ensures that Shusha's revival is not merely a reconstruction effort, but a living example of how technology and tradition can merge to create a resilient, intelligent, and culturally vibrant city.

## 8. Fuzuli and Jabrayil – Smart Urban Connectivity

### 8.1 Smart Mobility and Transport Infrastructure

The reconstruction of Fuzuli and Jabrayil embodies Azerbaijan's vision of intelligent connectivity, where smart mobility systems form the backbone of urban development. In Fuzuli, the creation of electric public transportation networks, autonomous shuttles, and integrated bike lanes represents a shift toward low-emission mobility solutions. The Fuzuli International Airport, one of the first large-scale projects in the liberated territories, serves as a cornerstone of this transformation. Equipped with AI-based navigation systems, digital passenger flow management, and energy-efficient lighting technologies, the airport demonstrates how smart transport infrastructure can operate sustainably while ensuring safety and convenience. These systems are connected through centralized data platforms that allow real-time monitoring of traffic density, vehicle emissions, and energy use, providing valuable insights for future urban planning.

### 8.2 Digital Infrastructure and Smart City Integration

Beyond mobility, the cities of Fuzuli and Jabrayil are being developed as digitally connected smart hubs where architecture, energy systems, and communication networks function as one integrated ecosystem. Smart grids are being established to distribute renewable energy efficiently, while IoT-enabled utilities monitor electricity, water, and heating consumption to reduce waste and costs. Urban planners utilize Building Information Modeling (BIM) and Geographic Information Systems (GIS) to ensure that infrastructure development follows a data-driven approach, minimizing spatial conflicts and optimizing land use. In Jabrayil, digital connectivity extends to public services—smart street lighting, sensor-based waste collection, and e-governance tools simplify management and enhance citizen interaction with the city's digital environment. Together, these systems reflect a broader strategy where digital architecture supports urban resilience and adaptive governance.

### 8.3 Environmental Resilience and Sustainable Growth

The vision for Fuzuli and Jabrayil also places strong emphasis on environmental sustainability. Urban expansion is guided by ecological mapping that identifies protected areas and green corridors to preserve biodiversity. Both cities adopt smart water recycling systems, solar-powered infrastructure, and green building materials that contribute to the reduction of carbon emissions. Advanced environmental sensors monitor air and soil quality, while green zones are strategically distributed to improve urban microclimates. These innovations ensure that technological growth in the region aligns with ecological principles, positioning Fuzuli and Jabrayil as pioneering examples of how digital connectivity and environmental consciousness can coexist in post-conflict reconstruction.

**9. Lachin and Kalbajar – Green Energy and Smart Resource Management**

**9.1 Renewable Energy Integration and Architectural Innovation**

The mountainous districts of Lachin and Kalbajar represent a central component of Azerbaijan's "Green Energy Zone" initiative, where digital architecture and renewable technologies merge to achieve environmental resilience and energy independence. The unique topography of these regions provides ideal conditions for hydropower generation, solar photovoltaic installations, and geothermal heating systems that are directly embedded into building designs. Local architectural projects prioritize passive energy efficiency, utilizing orientation, natural ventilation, and high-performance insulation materials to minimize energy demand. Smart meters and IoT-based monitoring systems allow for real-time tracking of electricity consumption, ensuring that each structure contributes to a collective energy balance within the region. This combination of digital design and renewable energy infrastructure positions Lachin and Kalbajar as benchmarks for sustainable mountain urbanism.

9.2 Smart Resource Management and Environmental Monitoring

Effective resource management is essential for maintaining ecological stability in fragile mountain ecosystems. In Lachin and Kalbajar, Geographic Information Systems (GIS) play a key role in mapping water resources, soil fertility, and vegetation coverage to guide development decisions. Through the use of remote sensing and environmental sensors, authorities can monitor biodiversity, detect deforestation risks, and control land degradation in real time. Smart water networks with automated leak detection and adaptive distribution models help to optimize the use of freshwater sources, while digital waste management systems reduce pollution. The integration of these digital tools not only safeguards natural resources but also promotes transparent and data-driven governance, aligning reconstruction with the broader principles of sustainable regional planning.

**9.3 Community Resilience and Sustainable Livelihoods**

Beyond infrastructure and technology, the transition to a smart and green energy model in Lachin and Kalbajar has a profound socio-economic dimension. Local communities are being trained in the operation of renewable energy systems and

sustainable construction practices, fostering long-term self-sufficiency. The development of eco-tourism and smart agriculture—supported by digital irrigation, precision farming, and green hospitality facilities—creates new economic opportunities while maintaining harmony with the natural landscape. By combining environmental protection, digital innovation, and social empowerment, Lachin and Kalbajar are becoming living examples of how sustainability and technology can coexist to shape resilient post-conflict regions.

## 10. Toward a Sustainable Smart Region

### 10.1 Integration of Smart Infrastructure

The reconstruction of Karabakh marks a transition from fragmented urban renewal to a unified, intelligent regional ecosystem. Smart infrastructure — from digital transport management to integrated energy grids — connects cities and villages into a single operational network. This system allows data sharing across urban zones, optimizing mobility, utilities, and public services. Through interoperability, each settlement contributes to the resilience of the entire region, ensuring that development remains balanced and resource-efficient.

### 10.2 Cultural Identity through Technological Design

Digital technologies in architecture and planning are not limited to efficiency; they become instruments of cultural restoration. Using Building Information Modeling (BIM) and digital twin simulations, architects integrate traditional Azerbaijani design principles into modern smart structures. Reconstructed public spaces reflect local heritage through materials, forms, and community-centered layouts. This fusion of technology and culture turns Karabakh's smart transformation into a symbol of renewal and identity continuity.

### 10.3 Regional Vision and Global Impact

As smart cities in Karabakh mature into interconnected urban ecosystems, they set a precedent for sustainable regional governance in Azerbaijan and beyond. The region evolves into a living laboratory for green innovation — showcasing how renewable energy, AI-based management, and digital participation can coexist with social inclusivity and cultural integrity. Karabakh's model demonstrates that smart reconstruction can transcend physical rebuilding — becoming a pathway toward a sustainable, data-driven, and human-centered future.

## References

1. Aghali Smart Village Project. (2022). *A Model of Smart Rural Development in the Post-Conflict Region*. Ministry of Economy of Azerbaijan.
2. Aghali Smart Village Project. (2022). *Digital and Sustainable Rural Development Model in Karabakh*. Official Government Report.

3. Alawadhi, S., Scholl, H. J., & Susha, I. (2012). *Building Smart Cities: Lessons from World Cities*. Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS).
4. Aliyeva, S. (2024). *Cultural Heritage Meets Technology: The Smart Restoration of Shusha*. *Journal of Sustainable Architecture*, 9(1), 71–88.
5. Azerbaijan Ministry of Digital Development and Transport. (2023). *Smart City and Smart Village National Strategy 2022–2026*. Baku.
6. Azerbaijan Ministry of Economy. (2022). *The Smart Village Concept in the Liberated Territories*. Baku: Ministry of Economy of the Republic of Azerbaijan.
7. Batty, M. (2018). *Inventing Future Cities*. MIT Press.
8. Cabinet of Ministers of the Republic of Azerbaijan. (2022). *Karabakh Green Energy Zone Declaration*. Official Policy Document.
9. Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). *Smart Cities in Europe*. *Journal of Urban Technology*, 18(2), 65–82.
10. Giffinger, R., & Gudrun, H. (2010). *Smart Cities Ranking: An Effective Instrument for the Positioning of the Cities? ACE: Architecture, City and Environment*, 4(12), 7–26.
11. Hasanov, E., & Huseynova, N. (2023). *Smart Reconstruction Strategies in Karabakh: Integration of Digital and Ecological Design*. *Azerbaijan Architecture and Construction Journal*, 11(2), 45–59.
12. Hollands, R. G. (2008). *Will the Real Smart City Please Stand Up? City: Analysis of Urban Trends, Culture, Theory, Policy, Action*, 12(3), 303–320.
13. Jabrayil District Development Office. (2024). *Smart Grid and Waste Management Systems in Reconstructed Areas*. Baku.
14. Karabakh Renaissance Foundation. (2024). *Shusha – The Smart Cultural Capital of Azerbaijan*. Baku.
15. Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications.
16. PMD Projects. (2024). *Sustainable Construction and Smart Development in Azerbaijan*. Baku: PMD Sustainability Report.
17. PMD Projects. (2024). *Sustainable Construction Practices in Azerbaijan: From Smart Materials to Green Certification*. Internal Report.
18. State Committee for Urban Planning and Architecture of Azerbaijan. (2022). *Digital Master Planning and Reconstruction Projects in Karabakh*. Baku.
19. State Committee on Urban Planning and Architecture of the Republic of Azerbaijan. (2023). *Master Plans for the Reconstruction of Karabakh and Eastern Zangazur*. Baku.
20. Townsend, A. M. (2013). *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. W. W. Norton & Company.
21. UNDP Azerbaijan. (2023). *Smart and Green Reconstruction of Karabakh: Building a Sustainable Future*. Baku: UNDP Country Office.

22. United Nations Economic Commission for Europe (UNECE). (2023). *Digital Transformation and Smart Territories in Eastern Europe and the Caucasus*. Geneva.
23. United Nations Human Settlements Programme (UN-Habitat). (2020). *People-Centered Smart Cities: A New Digital Agenda for Urban Development*. UN-Habitat Report.
24. World Bank. (2021). *Smart Cities and Sustainable Urban Development*. Washington, D.C.: World Bank Group.
25. World Bank. (2022). *Post-Conflict Urban Regeneration: Smart and Sustainable Development in South Caucasus*. Washington, D.C.

# A New Approach in Structural Control: An Industrial Application for the Tallest Building in the Caucasus

Azer Kasimzade[1[0000-0002-2487-3813]], Varol Koç [2[0000-0003-4810-3845]], Furkan Gunday [3[0000-0003-2979-9373]], Hasan Yılmaz [4[0009 0007 1961 1740]], Emin Nematli [1[0000-0003-2979-9373]]

[1] Azerbaijan University of Architecture and Construction, Baku, Azerbaijan
[2] Ondokuz Mayıs University, Samsun, Turkey
[3] Giresun University, Giresun, Turkey
[4] Milli Savunma Bakanlığı, Ankara, Turkey

azer.qasimzade@azmiu.edu.az

**Abstract.** For the Baku Tower (BT), the tallest building in the Caucasus region, constructed in Baku, the capital of Azerbaijan, controlling dynamic behavior is critical for both safety and user comfort. The structure is equipped with a Tuned Mass Damper (TMD) system with a mass of 400 tons to reduce vibrations caused by wind and earthquakes. In this study, the dynamic analysis results of the BT with and without the damper were compared using the Finite Element Method (FEM). The results show that the use of TMDs increases the period.

In the second phase of the study, the Structural Health Monitoring (SHM) system installed on the building is examined. Through accelerometers and sensors placed on critical floors, the dynamic behavior of the structure is continuously monitored, and warning signals are sent to the building management and power control systems when critical threshold values are exceeded, based on real-time comparisons of the obtained data. Therefore, the structural control system performance and the indicators obtained from numerical dynamic analyses are correlated with the real-time SHM system. Thus, the analysis results are evaluated beyond their use in design and verification, and integrated into building operation and safety decision-making processes. Critical parameters such as story displacements, vibration characteristics, and peak displacements obtained from dynamic analyses are used as monitorable threshold values by the SHM system. In this way, this data is transformed into operational decision criteria that can be transferred to the building control system. The integrated use of structural control systems and SHM applications in a high-rise building has been examined through analyses and applications.

**Keywords:** Structural control, Tuned mass damper (TMD), Structural health monitoring (SHM), High-rise buildings, Wind effect, Seismic behavior, Dynamic analysis, Vibration control, Decision support system, Smart building systems

# 1 First Section

In this study, the Baku Tower (BT) building was subjected to nonlinear and time-date interval dynamic analyses using finite element methods in both its normal and Tuned Mass Damper (TMD) system configurations, and its response values under free and forced vibrations were compared. The critical values obtained from the TMD system configuration were evaluated by a decision support system designed to determine the necessary maintenance or intervention for the structure. This system generates a warning signal when it detects that critical threshold values have been reached due to the stress of dynamic factors such as wind and earthquakes. The information is sent to the building's power control system via a transmission line, enabling the implementation of emergency safety procedures in the water, electricity, and natural gas systems. A Structural Health Monitoring (SHM) supported decision system was used in the building. SHM is a system that continuously monitors the dynamic effects of earthquakes and wind; vibrations caused by various devices or vehicles that may affect the structure inside or around it; and long-term damage developments that may occur in the structure due to ground or load effects, as well as the structural behavior under these factors, using sensors. In SHM, data from accelerometers positioned at critical points in the building are continuously monitored by real-time analysis software. This allows for tracking the long-term damage development of the structure. Whether the building has sustained damage after an earthquake can be quickly determined, enabling decisions to be made regarding the safe use of the building.

In this study, stages that are often carried out independently between analysis, monitoring, and building management processes are integrated into a single decision support chain. Therefore, structural performance evaluation is viewed not only as a process used in the design-review phase, but also as an active component in the continuous evaluation of the building's comfort of use and operational safety. Thus, the integrated use of structural control systems and smart building management applications in a high-rise building is examined through analysis and applications, and structural control systems, smart building management technologies, and smart building management applications are made usable in an integrated manner in engineering practice.

Currently, intensive research is being conducted on transferring data obtained from structural analysis (SAM) to building operation processes (Kamariotis et al., 2023; 2024). However, studies focusing on integrated approaches to integrating performance limits obtained from structural analysis results into building control and operation systems through real-time evaluation via the SAM system should be increased. While there are studies on performance-based SAM and control systems in the existing literature (Spencer and Nagarajaiah, 2003; Hughes et al., 2021), approaches for automatically transferring performance thresholds obtained from analysis to control systems using real-time SAM data are lacking. With the method proposed in this study, performance thresholds determined from analyses are defined in the SAM system, and sensor data are compared with the threshold values. The building safety status is automatically determined, and control systems are activated when necessary. In other words, the analysis, monitoring, decision, and control chain is integrated. Performance-based analysis during the design phase (Moehle and Deierlein, 2004; FEMA P-58, 2012), which is

discussed separately in the literature, and vibration reduction studies carried out with SHM and control systems during the service life (Connor, 2003; Kareem, 1999) are used together as a chain. By addressing the gap in the real-time evaluation of performance thresholds obtained from analyses via the SHM system and their transfer to the control system, the analysis results are not only kept at the design stage but are also transferred to the SHM system, compared with real-time measurements, and subsequently the structural performance level is determined. This allows for the automatic evaluation of the structural condition and, when necessary, the activation of control systems to generate control, warning, and evacuation decisions. In the literature, modal analysis-based SHM studies aim to detect damage through frequency and mode changes (Wang et al. 2008; Khan et al., 2020), generate damage probabilities using vibration measurements (Hughes et al., 2021; Del Pozo et al., 2023), but do not link these to performance levels. Data-driven SHM approaches, on the other hand, aim to detect anomalies in the structure (Worden et al., 2007; Zacharakis and Giagopoulos, 2022), generate damage alarms using sensor data (Wang and Ke, 2024; Li et al., 2024), but maintain a weak relationship with analyses.

Numerous studies in the current literature on reducing vibrations caused by wind and earthquake effects in high-rise buildings mostly focus on the design and performance of TMD or seismic isolation systems and various passive, semi-active, or active damping methods (Housner et al., 1997; Kelly, 1993; Soong and Dargusg, 1997; Kareem, 1999; Spencer and Nagarajaiah, 2003; Connor, 2003). Similarly, research on SHM systems focuses on issues such as sensor placement, vibration-based damage detection, determination of modal parameters, and rapid post-earthquake damage assessment (Doebling et al., 1996; Sohn et al., 2003; Li et al., 2004; Worden et al., 2007; Farrar and Worden, 2012). There is a need for studies that intersect these two areas, which are generally considered independently. Comprehensive studies should be conducted that combine research on the performance of structural control systems examined during the design phase with research on SHM systems used as monitoring and reporting tools. In particular, there is a significant gap in the literature regarding the conversion of the performance of vibration control systems used in high-rise buildings under real operating conditions into decision support mechanisms. This study includes an application for performance-based real-time building management in high-rise buildings by integrating analysis, monitoring, and operation processes.

## 2 Problem Statement

Traditional SHM (Survey Management) applications involve a system where sensor data is interpreted by experts, leading to subjective decision-making risks and slow post-disaster decision delays. Therefore, it is crucial to generate automatic damage indicators by detecting threshold exceedances from SHM data to create alarms and warnings. In a typical automated decision-making process, feature extraction from sensor data is used to perform threshold checks and generate alarms, but usually only a decision of whether damage exists or not is made (Farrar and Worden, 2012; Worden and Manson, 2007; Ni et al., 2009; Sun et al., 2023; Riveros et al., 2014). More advanced

systems attempt to determine whether the structure can continue to be used after damage detection. For this purpose, the probability of a specific measurement corresponding to damage is calculated. The risk level is determined as the product of this damage probability and the severity of the damage. The results are evaluated in terms of loss of use, economic loss, and safety risk. Decision levels are generally classified as safe use, restricted use, and evacuation requirement (Sohn et al., 2003; Farrar and Worden, 2012; Lynch and Loh, 2006). Most systems base these decisions on measurement statistics rather than analytical data. Therefore, current studies adopt a probabilistic decision approach. Instead of deterministic thresholds, Bayesian updates, probabilistic damage models, and reliability indices are used. In these approaches, the structural performance is expressed as a probability distribution, the distribution is updated with sensor data, and risk levels are recalculated (Hughes et al., 2021; Feng et al., 2021; Jia and Papadimitriou, 2025; Straub and Der Kiureghian, 2010). However, the models are often complex and their real-time application is limited. The fact that SHM measurements make risk predictions by generating damage indicators means that the decision-making process is not related to structural analyses. In other words, threshold values are mostly empirical or statistical. They do not depend on the actual analytical performance of the structure. Decision-making mechanisms and control processes, developed by comparing performance thresholds obtained from structural analyses with SHM measurements, ensure that thresholds are grounded in analyses derived from structural modeling, that performance-based decisions determine performance levels rather than just alarms, and that control systems can be triggered by decision outputs directly transferred to building systems. Currently, SHM systems detect damage but lack performance-based decision-making and control integration (Kamariotis et al., 2023; 2024; Limongelli, 2020).

While SHM systems classically relied on physics-based analyses, data-driven approaches have come to the forefront in the last 20 years due to the increasing amount of data. This transformation is seen in the application areas of damage detection, anomaly detection, and prediction/prognosis (Farrar and Worden, 2012; Worden et al., 2007; Farrar and Sohn, 2000; Sikorska et al., 2011). In this context, machine learning approaches such as artificial neural networks, support vector machines, random forests, k-nearest neighbors, and deep learning models are used to classify structural condition by extracting features such as frequency variations, acceleration records, mode shape differences, and vibration characteristics from sensor data (Worden and Manson, 2007; Farrar and Worden, 2012; Avci et al., 2021). These systems operate on the principle of creating damage classes by using sensor data for feature extraction and model training. The fundamental problem with these methods, which allow learning from data even if the damage mechanism is not explicitly modeled, is that the model only works as well as the training data and may not always correctly interpret new situations (Worden et al., 2007; Farrar and Worden, 2012; Avci et al., 2021).

In many cases, since damage data is unavailable, the system learns normal behavior using techniques such as autoencoder networks, Gaussian mixture models, PCA-based methods, isolation forests, and clustering algorithms, and identifies deviations from this behavior as anomalies (Farrar and Sohn, 2000; Chandola et al., 2009; Avci et al., 2021). Another problem here is that an anomaly does not always mean damage. Factors such

as temperature changes, wind conditions, and changes in land use can also create anomalies (Peeters and De Roeck, 2001; Sohn, 2007; Cornwell et al., 1999; Ni et al., 2012). Further studies on model-based prediction (prognosis) methods not only detect damage but also make inferences about how the damage will develop in the future. Farrar and Lieven (2007), Eleftheroglou et al., 2018; Li and Noman (2025) state that at this point, techniques such as time series prediction models, LSTM and RNN networks, Kalman filters, and Bayesian update methods are used. These approaches produce outputs such as remaining service life estimation, performance degradation prediction, and maintenance planning (Yang et al., 2022; Zhang et al., 2021; Zhang et al., 2024; Kamariotis et al., 2023; Lai et al., 2024). However, the fact that the prediction is mostly not related to physical performance levels leads to problems in AI-machine learning based SHM studies, such as the inability to link the results to engineering performance criteria, relate them to analysis-based thresholds, and integrate them into control systems, especially in the damage detection, anomaly identification, and prediction generation phases. In other words, the algorithm learns from the data but cannot produce engineering decisions. At this point, expert review becomes necessary (Luckey et al., 2021; Jia and Li, 2023; Plevris, 2024). This study aims to complete the decision and control chain by combining analysis-based performance thresholds with SHM and data-driven approaches.

Today, modern cities are evaluated not only in terms of their physical characteristics but also in terms of their digital infrastructure. Thus, decision-making mechanisms based on real-time data are created in areas such as energy management, transportation, security, infrastructure monitoring, and disaster management throughout the city (Batty et al., 2012; Zanella et al., 2014; Bibri and Krogstie, 2017; Kitchin, 2014). Indeed, the goal of the smart city approach is not only energy and transportation optimization. One of the critical goals is to create cities that are resilient to disasters and capable of rapid recovery. Therefore, efforts are being made to use SHM systems not only as individual building monitoring tools but also as an integral part of city management (Lu et al., 2025; Zhang et al., 2026; Bertino et al., 2021; Aroquipa et al., 2025; Rodríguez-Antuñano et al., 2022). However, it is necessary to advance studies on integrating the information obtained from SHM (Strategic Engineering Monitoring) data of structures considered as passive components in smart city systems into the smart city management mechanism. In particular, high-rise buildings, bridges and viaducts, metro systems, hospitals, critical public buildings, energy and communication infrastructures should be monitored as parts of a network system. Thanks to this integration, damaged and at-risk areas, as well as inoperable infrastructure, can be seen instantly. However, in most applications, only damage maps are produced, and performance evaluation remains limited (Bertino et al., 2021; Jahanshahi and Masri, 2012). The most critical problem after an earthquake is the rapid identification of buildings that are unsafe and should not be used, as well as safe areas that need to be evacuated (Applied Technology Council, 1991). The traditional method, where engineers go to the site to inspect buildings, causes the process to take weeks. Although speed has been achieved with SHM systems, most systems only provide warnings about potential damage. The actual performance level is not determined (Sohn et al., 2004; Worden and Farrar, 2007).

The goal of modern disaster management is to direct search and rescue teams to the right places, protect hospital and transportation infrastructure, and rapidly restore city functions. In current systems, building performance levels cannot be directly linked to city management decisions. While damage maps can be generated from post-disaster data, the inability to automatically determine performance levels, the lack of correlation between decision-making mechanisms and analysis, and the limited evaluation of existing data in decision-making indicate the shortcomings of current data. However, the use of performance thresholds obtained through analysis, especially when applied at the city level, automates and accelerates the identification of safe areas, the closure of buildings, and the optimization of emergency response plans. The potential of this study, extending from the building scale to the city scale, can provide a revolutionary improvement in multi-disaster management systems. In the traditional approach, performance evaluation is largely limited to the design and construction phases, and behavioral changes under various factors are identified through post-damage inspections. Thanks to advancements in sensor technologies, data processing capacity, and communication infrastructures, buildings should now be considered as systems that continuously generate data, and this data must be effectively utilized in smart city approaches. Just as structural engineering (SHM) systems become not merely a monitored parameter but an active component guiding operational decisions when integrated into smart building management, the integration of city-scale building data into central disaster and infrastructure management systems can create holistic emergency response and service sustainability. By integrating analysis, monitoring, and operation processes, chaotic post-disaster decision-making processes can be replaced by data-driven management strategies, leading to significant advantages in multi-disaster management. This study presents the building-scale engineering equivalent of data-driven management approaches developed for smart cities and offers a framework for the future integration of SHM-supported performance management with city-scale disaster and infrastructure management systems.

## 3 SHM and TMD Systems Installed in the Baku Tower Building

Baku, the capital of Azerbaijan, located on the Caspian Sea coast and affected by varying degrees of winds daily, derives its name from the word "bakı," meaning "windy place." This meaning alone highlights the challenges of constructing high-rise buildings in Baku. As the building gets higher, the increasing wind intensity causes vibrations and oscillations. Even low-acceleration oscillations of 10-20 milli-g, if sustained for a period, can be felt by people, creating a "seasickness" effect. Therefore, it is necessary to dampen the oscillations caused by wind in high-rise buildings for comfort. Especially in high-rise buildings like the Baku Tower (BT), which was constructed between 2014 and completed in 2020, and is the tallest building in the Caucasus region at 276.3 m, it is essential to prioritize comfort as well as safety and functionality, given the constant

and sometimes very strong wind effects. For this purpose, floor rigidities, damping ratios, and mass distributions must be optimized. BT, with its roof and observation platform, has 50 above-ground floors and 2 larger underground floors. As with all high-rise buildings, shear forces are high in the lower floors and displacements are high in the upper floors. To limit these displacements, a 400-ton Tuned Mass Damper (TMD) is used in BT to dampen wind and seismic effects. The mass block, consisting of a large steel box filled with in-situ concrete, undergoes controlled horizontal movement of 1.3 m in each direction due to wind and seismic effects. Thus, the TMD reduces accelerations from the lower limit of 10 milli-g to 7 milli-g. This reduces the dynamic loads acting on the structure and ensures living and working comfort in BT, which consists of office centers, commercial and service areas (Structurae, 2020).

Figure 1 shows a view from the top of the tower towards the interior compartments of the steel box before the concrete was poured. The reinforced concrete reinforcement is also visible at the bottom. The block has four suspension systems, each containing three cables, the housings of which are shown in Figure 1. Although four cables would normally suffice according to calculations, twelve cables were used to ensure the highest level of safety. These pendulum cables are manufactured as smooth-surfaced, fully interlocking cables. As can be seen in Figure 2, the cables are used in sets of three at each of the four suspension points. In this way, the cross-linked hydraulic dampers are also visible. Figure 3 shows the TMD system and some of its features. Views from the construction phase of the BT are given in Figure 4. The TMD system effectively produces benefits when its vibration period is aligned with the natural period of the structure. Therefore, natural frequency measurements had to be taken after the tower was completed. TMD adjustment was made through cables extended to coincide with the natural frequency. For this purpose, an adjustment block, which acts as a mechanical compression device, was used. This block, easily movable downwards within the steel box, allows for the extension of the free pendulum length at the top, thereby slowing down pendulum vibrations. Following vibration testing, a total of ten hydraulic dampers were installed, mounted diagonally around the mass block. These dampers ensure that pendulum deflection caused by hurricane and earthquake effects is limited to a maximum of 1.3 m. Since hydraulic dampers react according to speed, they show almost no reaction under effects that would cause relatively small pendulum movements, such as strong winds. TMD reduces large seismic effects by 10-15% overall. In addition, Anti-buckling Braces (BRBs) have also been installed on the tower as additional seismic protection (Structurae, 2020).

Both passive and adaptive TMD systems require a suitable and simple design to minimize costs and installation space. In this context, the slope of the oil dampers, the design of the TMD mass as simply as possible to minimize processing costs, cable anchors with variable compression units, and a robust suspension beam structure are critical. The secondary mass added to the structure with the TMD moves in phase opposite to the vibration of the main structure, damping some of the energy and thus reducing the acceleration acting on the main structure. Wind analyses performed for BT have shown that ISO 10137 comfort limits cannot be achieved without TMD (Weber et al., 2019). Since actual wind loads are broadband and a purely frequency-tuned, undamped

system is only effective at a single frequency, TMD systems with internal damping are required for high-rise structures under severe wind loads. In BT, with dampers placed at a 60° angle, excessive displacements are limited and the system occupies less space as the damping force increases proportionally with the load. However, controlling vibrations affecting structures today is not only a design problem aimed at reducing structural damage, but also a performance problem concerning user comfort and operational safety (Weber et al., 2019). Wind effects on high-rise buildings and bridges, human-induced vibrations on overpasses, and ground vibrations from vehicles and machinery can cause unacceptable accelerations in terms of user comfort, even if they do not exceed the structural strength limits. The modern engineering approach is based not only on preventing structural damage and collapse, but also on controlling service performance. Although the use of TMD and isolation systems serving these purposes has become widespread, the performance of these systems is generally considered at the design stage. However, the performance of these systems should also be continuously monitored during operation. Therefore, today, it is becoming increasingly important to convert performance thresholds obtained from structural analysis into operational decisions via SHM (Strategic Maintenance and Control).

**Fig. 1.** Internal compartments of the steel box (Structurae, 2020).

**Fig. 2.** Cross hydraulic dampers and pendulum cables.

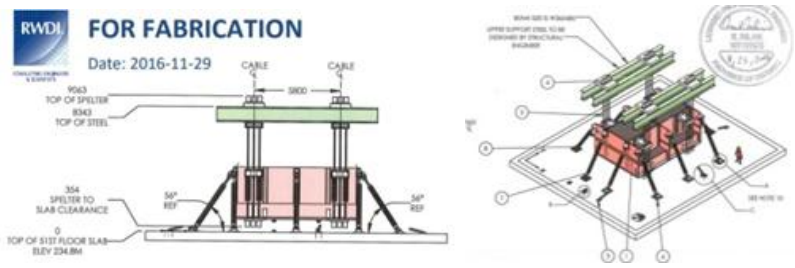



**Fig. 3.** TMD system.

The primary goal when establishing an SHM (Surface-to-Surface Hydraulic Works) system is to accurately capture the dynamic behavior of a building using the minimum number of accelerometers and sensors placed at critical floors that can represent this behavior. Initially, accelerometers are placed at the base of the foundation/basement, the first point where the actual seismic motion reflected from the ground to the building can be detected. This allows for a comparison of the ground motion with the upper floor movements, and the amplification effect of the structure can be determined from the measured difference. However, for this, sensors and accelerometers capable of measuring the maximum response of the structure are also placed at the top of the highest floor. This is because horizontal displacement, acceleration, and oscillation values are at their maximum at this point. Thus, the total oscillation of the building, the natural vibration period, and damping ratios can be obtained. Furthermore, the maximum wind speed the structure may be subjected to is also measured by a sensor placed at the top. In addition to all this, it is necessary to detect the different oscillation modes that occur, especially in high-rise buildings, due to their lack of a single, unified block movement under seismic influence. For this purpose, accelerometers should also be placed at critical middle floors. This allows us to obtain information such as the distribution of motion between floors, vibration modes, and where damage is concentrated or may be concentrated. In soft stories, transfer floors, floors where the floor height changes, or where shear walls are reduced, the change in stiffness increases the floor displacement, thus increasing the risk of damage. Torsional motion can also be captured by placing more than one accelerometer on some floors. The aim of SHM is not to place sensors and accelerometers on every floor, but to place sensors and accelerometers at the minimum number of points that can represent the dynamic behavior of the structure.

In addition to accelerometers, a wind sensor has also been installed at the top of the structure. Thanks to both 2-axis and 1-axis accelerometers placed at the mezzanine floors, displacements in directional orientation and torsional effects can be detected. Examples of the application and planning stages of the accelerometer system installation are presented in Figure 6.

In the study by Sevgen et al. (2025), which describes the planning phase of the SHM system installation in the IT building, the control of power systems is not addressed. This is because the study describes the planning required for the SHM system to send warning signals to the control unit when critical threshold values are detected. It is assumed that the necessary subsequent actions will be performed by the power control

system according to the specified protocol in emergency mode. In smart building systems, in any disaster or emergency, the building's power system control unit performs many defined tasks related to building safety depending on the type of warning. For these actions to be performed effectively, an accurate and reliable warning signal must be transmitted to the power control unit, and the control unit must be properly informed. At this point, this study includes analyses aimed at analytically determining critical threshold values.

**Fig. 4.** Views of the IT construction phases (Council on Vertical Urbanism, n.d.)

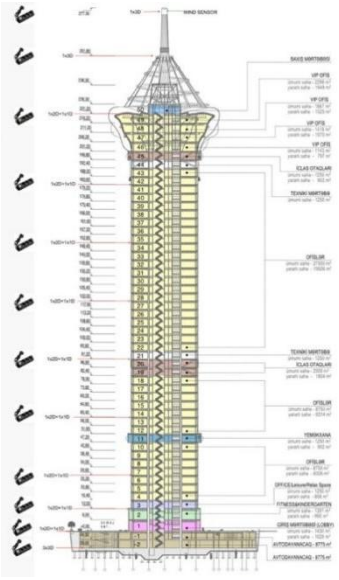


**Fig. 5.** General and usable area characteristics of floors in the CT system and location characteristics and number of accelerometers used for SHM.

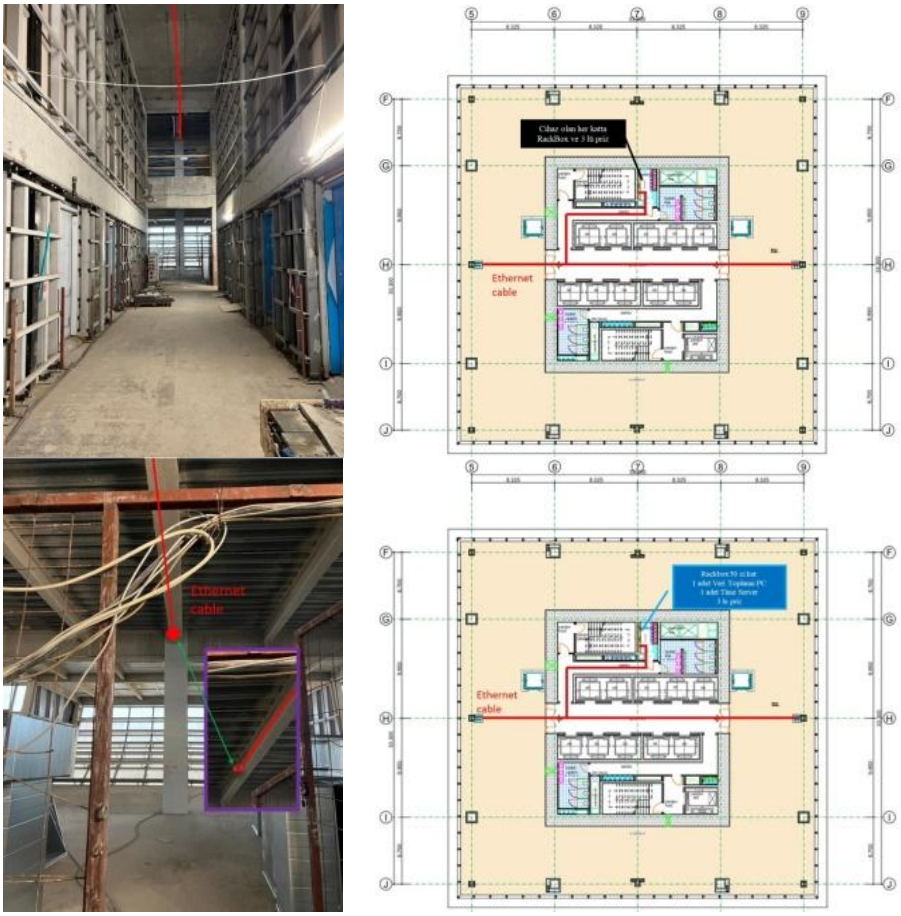



**Fig. 6.** Examples of implementation and planning for accelerometer placements on some floors.

## 4 Method

In this study, the method proposed is to determine the alarm thresholds to be used in the SHM system based on the peak acceleration, peak displacement, and maximum story displacement values obtained from time-domain nonlinear analyses performed on the numerical model of the structure. This study essentially presents a methodological proposal. Therefore, only two earthquake records were used. For more comprehensive studies, calibration can be performed with more earthquake records. In this study, Chi-Chi and Kobe earthquake records were selected as reference earthquake records due to their different frequency contents and strong ground motion characteristics, and the structural responses obtained under these records were evaluated as threshold values

representing the reference demand levels for the system. For this purpose, modal analyses were performed for the damped and undamped cases of the BT building using the Etabs V9.7.4 FEM program. In the modeling of the TMD-equipped BT structure with ETABS, the structure mass was taken as 80,688,774.55 kg, the TMD mass as 400,000 kg, and the structure mass/TMD mass = 201.6, using the IBC 2012 Regulation. The dead load of the viscous damper, steel beam, and other elements used to fix the TMD's Mass Box was taken as 80,771.8 kg. Antenna mass of 156,000 kg was added. The total length and width of the cover on which the loads rest was taken as 16,625 x 19.80 m; the total area was taken as 329,175 $m^2$.

For damper case analyses, the lateral stiffness (k) of the TMD system was taken as 544.32 kN/m and the period as 5.3835 sec. To obtain the design-based behavior spectrum, acceleration spectra (uniform hazard spectra) obtained based on soil-dependent spectral amplitudes at periods Ss= 0.2s and S1= 1s in the IBC (2012) seismic code were used. In this stage, the spectral acceleration modification coefficients included in the code were used to calculate the effect of local soil conditions on spectral accelerations. A D-type soil class was assumed for the local soil conditions in the project area. Therefore, there is a modification in the spectral accelerations obtained under the reference soil conditions. The spectral accelerations calculated according to the periods are given in Figure 7. Based on these calculations, the SS=1.18 and S1=0.7 information specified in the BT Calculation Report were entered into the Seismomatch software.

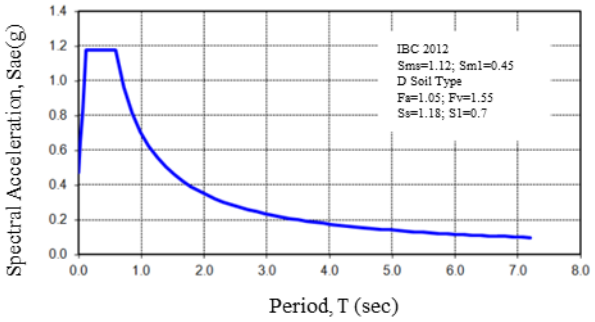



**Fig. 7.** Period- Spectral Acceleration Graphic

## 5 Findings

According to the analysis results, the period values for the modes are given in Table 1, and the peak horizontal displacement values are given in Table 2. The period value for the first mode was obtained as 4.393 seconds in the undamped case and 4.705 seconds in the damped case. The period value for the second mode was obtained as 4.023 seconds in the undamped case and 4.3 seconds in the damped case. Thus, it was observed that the period of the structure could be increased by 7.1% in the first mode and 6.88% in the second mode with the TMD system. The peak horizontal displacements

were 0.193 m in the undamped case and 0.181 m in the damped case, and it was observed that the peak displacements could be reduced by 6.17% thanks to the TMD system (Table 2). The comparison of the forced vibration responses of damped and undamped BT structures under the Kobe earthquake is given in Table 3. The comparison of the forced vibration responses of the damped and undamped BT structure under the Chi-Chi earthquake is given in Table 4. Forced vibration ETABS BT analysis results are given in Figures 8-12 for the Kobe earthquake and in Figures 13-17 for the Chi-Chi earthquake. The maximum drift in the Kobe EQ with damper, story 37 - drift in the X direction was obtained as 0.005199 m; the maximum drift in the Kobe EQ without damper, story 41 - drift in the X direction was obtained as 0.007977 m. This shows that a 35% reduction in maximum drift was achieved in the damped system. The maximum drift in the Chi-Chi EQ with damper, story 37 - drift in the X direction was obtained as 0.002808 m; the maximum drift in the Chi-Chi EQ without damper, story 42 - drift in the X direction was obtained as 0.004424 m. This shows that a 36% reduction in maximum drift is achieved in the damper system.

**Table 1.** Comparison of Free Vibration Responses for BT with damper and BT without damper structure

| Mode | BT Period [sec] Without TMD | BT Period [sec] With TMD | Difference [%] |
|---|---|---|---|
| 1 | 4.3721 | 5.4946 | 25.67 |
| 2 | 4.0033 | 5.4596 | 36.38 |
| 3 | 1.3844 | 4.3362 | 213.22 |
| 4 | 1.1256 | 3.9878 | 254.28 |
| 5 | 1.0858 | 1.3637 | 25.59 |
| 6 | 1.0088 | 1.1400 | 13.01 |
| 7 | 0.9722 | 1.1123 | 14.41 |
| 8 | 0.9424 | 1.1053 | 17.29 |
| 9 | 0.4486 | 1.0889 | 142.73 |
| 10 | 0.3993 | 1.0725 | 168.60 |
| 11 | 0.2382 | 1.0051 | 321.96 |
| 12 | 0.2088 | 0.9967 | 377.35 |

**Table 2.** Peak horizontal displacement values.

| Peak point displacement [m] X direction | | Difference [%] |
|---|---|---|
| Without TMD | With TMD | |
| 0.19339 | 0.18145 | -6.174 |

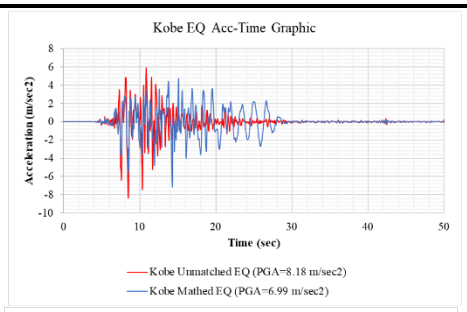
**Table 3.** Comparison of Forced Vibration Responses for BT with damper and BT without damper structure under the Kobe Earthquake

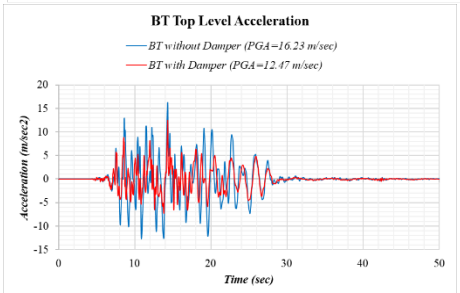
| | Without TMD | With TMD | Difference [%] |
|---|---|---|---|
| Acc (Top) | 1.623e+01 | 1.247e+01 | -23.17 |
| Disp (Top) | 9.788e-01 | 8.778e-01 | -10.31 |

| | | | |
|---|---|---|---|
| Base Shear | 3.204e+08 | 8.130e+07 | -74.63 |
| Base Mom | 1.877e+10 | 1.725e+10 | -8.1 |

**Table 4.** Comparison of Forced Vibration Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake

| | Without TMD | With TMD | Difference [%] |
|---|---|---|---|
| Acc (Top) | 8.544e+00 | 4.692e+00 | -45.08 |
| Disp (Top) | 5.011e-01 | 4.699e-01 | -6.23 |
| Base Shear | 1.996e+08 | 2.797e+07 | -85.99 |
| Base Mom | 1.112e+10 | 8.410e+09 | -24.37 |



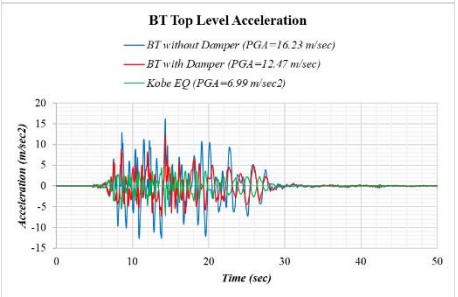
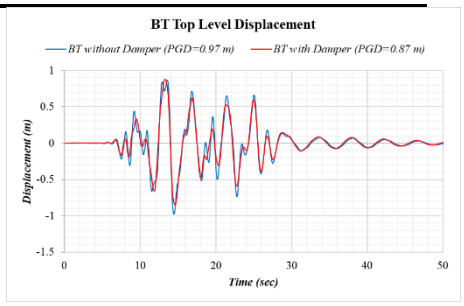
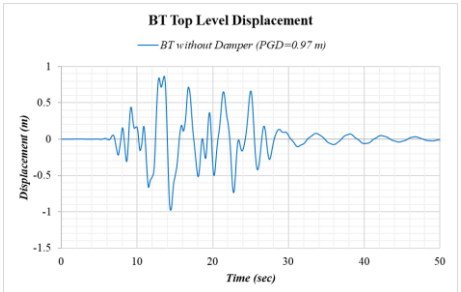




**Fig. 8.** Kobe EQ Acc-Time Graphic, Comparison of Top-Level Acceleration Responses for BT with damper and BT without damper structure under the Kobe Earthquake



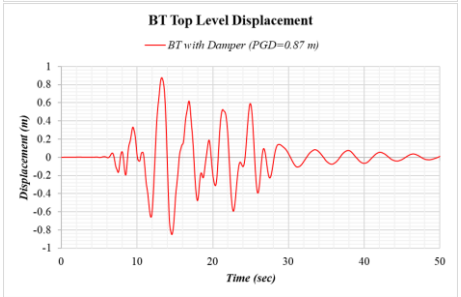




**Fig. 9.** Comparison of Top-Level Displacement Responses for BT with damper and BT without damper structure under the Kobe Earthquake

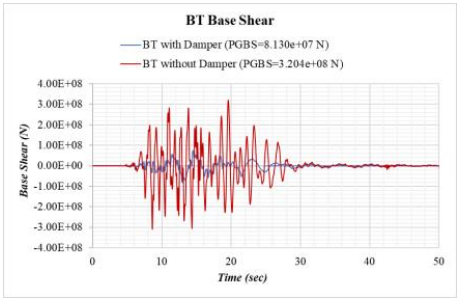



**Fig. 10.** Comparison of Base Shear Responses for BT with damper and BT without damper structure under the Kobe Earthquake

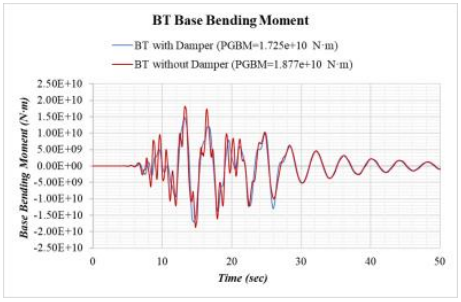



**Fig. 11.** Comparison of Base Bending Moment Responses for BT with damper and BT without damper structure under the Kobe Earthquake

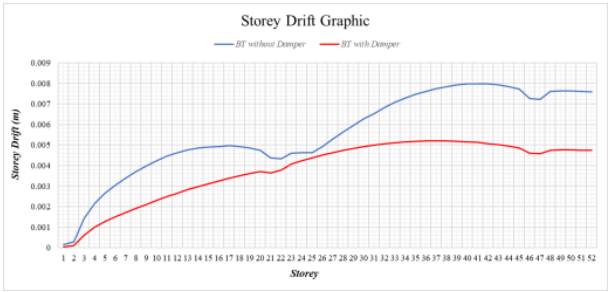
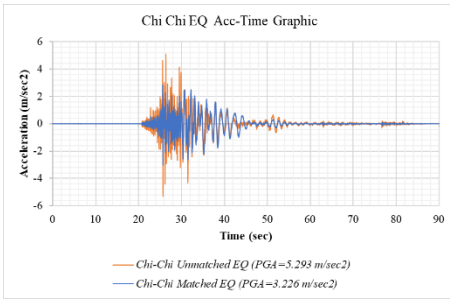
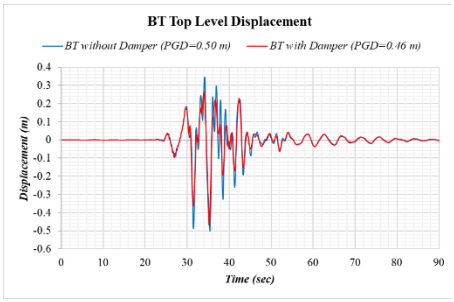



**Fig. 12.** Comparison of Storey Drift Responses for BT with damper and BT without damper structure under the Kobe Earthquake.



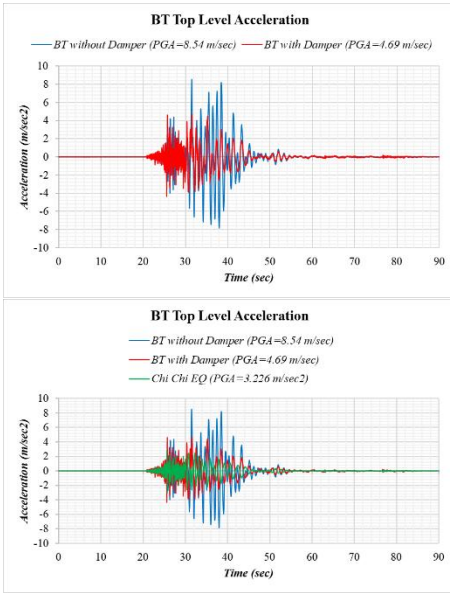



**Fig. 13.** Chi Chi EQ Acc-Time Graphic, Comparison of Top-Level Acceleration Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake

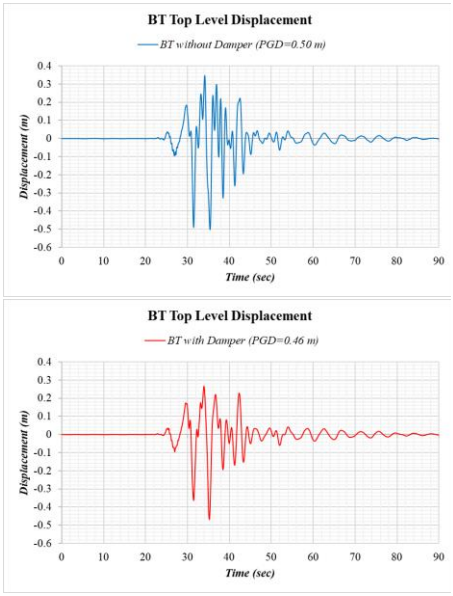



**Fig. 14.** Comparison of Top-Level Displacement Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake

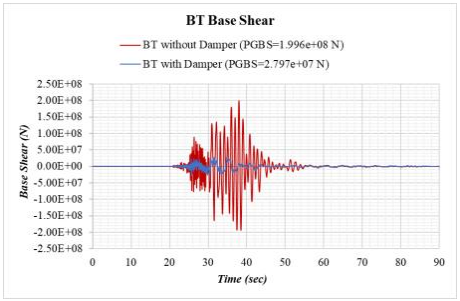



**Fig. 15.** Comparison of Base Shear Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake

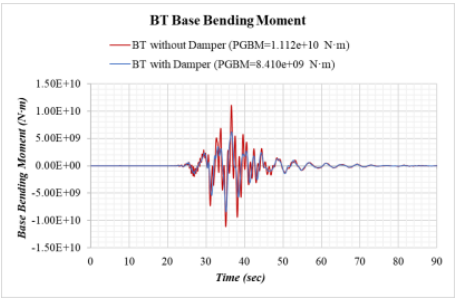



**Fig. 16.** Comparison of Base Bending Moment Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake

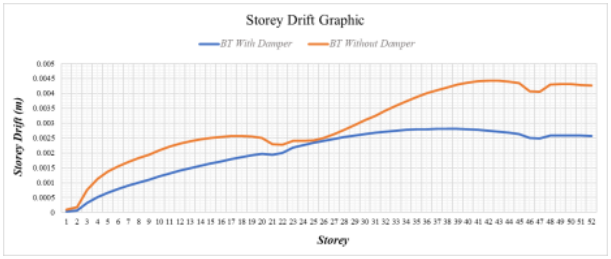



**Fig. 17.** Comparison of Storey Drift Responses for BT with damper and BT without damper structure under the Chi Chi Earthquake.

## 6 Comparing with SSIS Approach

The application of a 400-ton Tuned Mass Damper (TMD) system to mitigate wind and seismic effects in the Baku Tower (BT) has significantly improved the dynamic responses of the structure.

Furthermore, by applying the more modern and efficient Structural Seismic Isolation Method (SSIM) for the vibration control of highly reliable structures (HRS), a higher degree of reduction in structural responses can be ensured (Table 5).

By applying the SSIM method, the contact surface between the building base and the foundation is made spherical, and these surfaces are connected to each other with elastomeric devices (LRB, LCRB,). The resulting SSIS system thus gives the structure an inverse pendulum behavior.

Since the building rotates as a single rigid body (inverse pendulum) on the spherical foundation, story drifts are reduced to almost zero. This perfectly protects both the structure itself and the non-structural elements within it (e.g., glass facades, partitions, elevator shafts, etc.)."

**Table 5.** Comparison of reduction percentages in dynamic parameters provided by TMD (Baku Tower) and SSIS systems

| Dynamic Parameter | Reduction Provided by TMD System (BT - Kobe) | Average Reduction Provided by SSIS System for High-Rise Buildings [34] | Average Reduction Provided by SSIS System for Tower Structure [35, 36] |
|---|---|---|---|
| Story Drift | ~35 % | Extremely low due to rigid rotation | Extremely low due to rigid rotation |
| Acc (Top) | 23.17 % | 46.52 % | 82.43 % |
| Disp (Top) | 10.31 % | 18.75 % | 82.76 % |
| Base Shear | 74.63 % | 78.28 % | 96.59 % |
| Base Moment | 8.10 % | 86.98 % | 89.11 % |

## 7 Conclusions

Since the current IT building has dampers, it is possible to use the results obtained from the analyses performed for the damper condition as approximate threshold values. As stated in the findings section, Top Acc. was obtained as 4.69 in the Chi-Chi earthquake and 12.47 in the Kobe earthquake. In this case, when the Top Acc. values obtained from the SHM system are less than 4.69, it is a normal situation; when they are between 4.69 and 12.47, it is a warning situation; and when they are greater than 12.47, it indicates high demand followed by a critical situation. Top Disp. was obtained as 0.469 m in the Chi-Chi earthquake and 0.878 m in the Kobe earthquake. In this case, when the Top Disp. values obtained from the SHM system are less than 0.469 m, it is a normal situation; when they are between 0.469 m and 0.878 m, it indicates a warning situation; and when they are greater than 0.878 m, it indicates high demand followed by a critical situation. In the damper-equipped condition, the maximum drift was determined to be 0.002808 m for the Chi-Chi earthquake and 0.005199 m for the Kobe earthquake. In this case, drift values obtained from the SHM system indicate a normal condition when less than 0.0028, a warning condition when between 0.0028 and 0.0052, and a high demand followed by a critical condition when greater than 0.0052. In warning conditions, structural and system inspections are recommended; in high demand conditions, urgent engineering assessments; and in critical conditions, emergency evacuation and closure of the building are suggested. As stated in the findings section, the damper provides a 35-36% reduction in maximum drift in the structure. If the SHM data shows that the maximum drift values decrease over time from this percentage and approach the drift values of the damper-free condition, it can be understood that the damper performance is starting to decline and the damper system needs to be checked and adjusted.
With the application of the SSIM method, structural responses are reduced by an average of 82%-96%, and story displacements are minimized to the point of being negligible

## References

1. AApplied Technology Council, & California. Office of Emergency Services. (1991). Procedures for postearthquake safety evaluation of buildings (Vol. 20). Governor's Office of Emergency Services.
2. Aroquipa, H., Hurtado, A., Murga, C., De La Cruz, R., & Tarque, N. (2025). Towards Smart Cities: Foundational Methodology for Implementing Intelligent Circular Resilience in Heritage Buildings through Structural Health Monitoring and Digital-Twins–Part A. International Journal of Disaster Risk Reduction, 105749.
3. Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical systems and signal processing, 147, 107077.

4. Aytulun, E., & Soyöz, S. (2020). Monitoring the structural health of a high-rise building before, during, and after an earthquake. Turkish Journal of Earthquake Research. 2(1), 61–75. https://doi.org/10.46464/tdad.735239
5. Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., ... & Portugali, Y. (2012). Smart cities of the future. The european physical journal special topics, 214(1), 481-518.
6. Bertino, E., Jahanshahi, M. R., Singla, A., & Wu, R. T. (2021). Intelligent IoT systems for civil infrastructure health monitoring: a research roadmap. Discover Internet of Things, 1(1), 3.
7. Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. Sustainable cities and society, 31, 183-212.
8. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.
9. Chen, A. (2008). Design Research of a Digital Database for Historic Buildings in Baku. In ARCC Conference Repository. pp 45- 49.
10. Choi, H. S., Sze, J., Ihtiyar, O., & Joseph, L. (2014). Overview of Seismic Loads and Application of Local Code Provisions for Tall Buildings in Baku, Azerbaijan. International journal of high-rise buildings, 3(1), 65-71.
11. Connor, J. J. (2003). Structural motion control. Pearson Education, Inc New York, NY, USA, 770, 771.
12. Cornwell, P., Farrar, C. R., Doebling, S. W., & Sohn, H. (1999). Environmental variability of modal properties. Experimental techniques, 23(6), 45-48.
13. Council on Vertical Urbanism. (n.d.). Baku Tower. The Skyscraper Center. https://www.skyscrapercenter.com/building/baku-tower/16337
14. Del Pozo,G. A., Svendsen, B. T., & Øiseth, O. (2023). Implementation of decision analysis on a structural health monitoring system applied to a bridge benchmark study. Structural Health Monitoring.
15. Diclek, U., & Der Kiureghian, A. (2022). The value of seismic structural health monitoring for post-earthquake building evacuation. Bulletin of Earthquake Engineering, 20, 4367–4393. https://doi.org/10.1007/s10518-022-01375-2
16. Doebling, S. W., Farrar, C. R., Prime, M. B., & Shevitz, D. W. (1996). Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review.
17. Eleftheroglou, N., Zarouchas, D., Loutas, T., Alderliesten, R., & Benedictus, R. (2018). Structural health monitoring data fusion for in-situ life prognosis of composite structures. Reliability Engineering & System Safety, 178, 40-54.
18. Farrar, C. R., & Lieven, N. A. (2007). Damage prognosis: the future of structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 623-632.
19. Farrar, C. R., & Sohn, H. (2000). Pattern recognition for structural health monitoring. In Workshop on Mitigation of Earthquake Disaster by Advanced Technologies. MEDAT.
20. Farrar, C. R., & Worden, K. (2012). Structural health monitoring: a machine learning perspective. John Wiley & Sons.
21. FEMA P- 58 (2012). Seismic Performance Assessment of Buildings, Volume 1 – Methodology, Federal Emergency Management Agency, Washington, DC
22. Feng, Z., Lin, Y., Wang, W., Hua, X., & Chen, Z. (2020). Probabilistic updating of structural models for damage assessment using approximate bayesian computation. Sensors, 20(11), 3197.

23. Gümüş, O. (2021). Guidelines for structural health monitoring systems on tall buildings and a case study (Yüksek lisans tezi). Middle East Technical University. https://open.metu.edu.tr/handle/11511/89732
24. Housner, G., Bergman, L. A., Caughey, T. K., Chassiakos, A. G., Claus, R. O., Masri, S. F., ... & Yao, J. T. (1997). Structural control: past, present, and future. Journal of engineering mechanics, 123(9), 897-971.
25. Hughes, A. J., Barthorpe, R. J., Dervilis, N., Farrar, C. R., & Worden, K. (2021). A probabilistic risk-based decision framework for structural health monitoring (preprint). arXiv. https://arxiv.org/abs/2101.01521
26. Limongelli, M. P., & Çelebi, M. (Eds.). (2019). Seismic Structural Health Monitoring: From Theory to Successful Applications. Springer. https://doi.org/10.1007/978-3-030-13976-6
27. Hughes, A. J., Barthorpe, R. J., Dervilis, N., Farrar, C. R., & Worden, K. (2021). A probabilistic risk-based decision framework for structural health monitoring. Mechanical Systems and Signal Processing, 150, 107339.
28. Hughes, A. J., Barthorpe, R. J., Dervilis, N., Farrar, C. R., & Worden, K. (2021). A probabilistic risk-based decision framework for structural health monitoring. Mechanical Systems and Signal Processing, 150, 107339.
29. Jahanshahi, M. R., & Masri, S. F. (2012). Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures. Automation in Construction, 22, 567-576.
30. Jia, J., & Li, Y. (2023). Deep learning for structural health monitoring: Data, algorithms, applications, challenges, and trends. Sensors, 23(21), 8824.
31. Jia, X., & Papadimitriou, C. (2025) Reliability Updating using Monitoring Data. Available at SSRN 5591513.
32. Kamariotis, A., Chatzi, E., & Straub, D. (2023). A framework for quantifying the value of vibration-based structural health monitoring. Mechanical Systems and Signal Processing, 184, 109708.
33. Kamariotis, A., Chatzi, E., Straub, D., Dervilis, N., Goebel, K., Hughes, A. J., ... & Worden, K. (2024). Monitoring-supported value generation for managing structures and infrastructure systems. Data-Centric Engineering, 5, e27.
34. Kasimzade, Azer & Abrar, Obaidullah & Atmaca, Gencay & Kuruoglu, Mehmet. (2020). New structural seismic protection for high-rise building structures. Journal of Vibroengineering. 22. 831-848. 10.21595/jve.2019.20741.
35. Kasimzade AA, Nematli E (2023) New Structural Seismic Isolation System for Tower Structures, International Congress on Advanced Earthquake Resistance Structures (AERS2023), Baku, Azerbaijan, 26-29 April, 2023.
36. Kasimzade AA, at all, Structural Seismic Isolation Method for New Generation Seismic Protection Strategy, Journal of Applied and Computational Mechanics, University of Ahvaz, 2026.
37. Kareem, A. (1999). Mitigation of Motions of Tall Buildings with Specific Examples of Resent Applications. Wind & Structures, 2(3), 208-211.
38. Kelly, J. M. (1993). Earthquake-resistant design with rubber (Vol. 7). London: Springer-Verlag.
39. Khan, M. W., Akmal Din, N., & Ul Haq, R. (2020). Damage detection in a fixed-fixed beam using natural frequency changes. Vibroengineering Procedia, 30, 38-43.
40. Kitchin, R. (2014). The real-time city? Big data and smart urbanism. GeoJournal, 79(1), 1-14.

41. Korkmaz, A., & Kayhan, A. H. (2016). Comparison of displacement-based methods with elastic and time history dynamic analysis methods. Trakya Üniversitesi Fen Bilimleri Dergisi, 9(1), 7–13. https://dergipark.org.tr/tr/pub/trakyafbd/issue/22997/245970
42. Lai, L., Dong, Y., & Smyl, D. (2024). SHM-informed life-cycle intelligent maintenance of fatigue-sensitive detail using Bayesian forecasting and Markov decision process. Structural Health Monitoring, 23(1), 187-210.
43. Li, B., Li, Y., & Noman, K. (2025). Structural health monitoring: Advanced sensing, diagnostics and prognostics. Sensors, 25(5), 1313.
44. Li, H. N., Li, D. S., & Song, G. B. (2004). Recent applications of fiber optic sensors to health monitoring in civil engineering. Engineering structures, 26(11), 1647-1657.
45. Li, X., Zhang, F. L., Xiang, W., Liu, W. X., & Fu, S. J. (2024, December). Structural health monitoring system based on digital twins and real-time data-driven methods. In Structures (Vol. 70, p. 107739). Elsevier.
46. Liang, Y., Huang, L., & Yang, J. (2024). Automated operational modal analysis for supertall buildings based on a three-stage strategy and modified hierarchical clustering. Structures, 62, 106194. https://doi.org/10.1016/j.istruc.2024.106194
47. Limogelli, M. P. (2020). SHM for informed management of civil structures and infrastructure. Journal of Civil Structural Health Monitoring, 10(5), 739-741.
48. Lu, J., Liu, L., Lei, Y., & Wei, J. (2025). Assessing the Impact of Smart City Development on Urban Resilience: An Empirical Study Based on China's Smart City Pilot Policy. Sustainable Cities and Society, 107029.
49. Luckey, D., Fritz, H., Legatiuk, D., Peralta Abadía, J. J., Walther, C., & Smarsly, K. (2021). Explainable artificial intelligence to advance structural health monitoring. In Structural health monitoring based on data science techniques (pp. 331-346). Cham: Springer International Publishing.
50. Lynch, J. P., & Loh, K. J. (2006). A summary review of wireless sensors and sensor networks for structural health monitoring. Shock and vibration digest, 38(2), 91-130.
51. Mardanshahi, A., Sreekumar, A., Yang, X., Barman, S. K., & Chronopoulos, D. (2025). Sensing techniques for structural health monitoring: A state-of-the-art review on performance criteria and new-generation technologies. Sensors, 25(5), 1424. https://doi.org/10.3390/s25051424
52. Moehle, J., & Deierlein, G. G. (2004). A framework methodology for performance-based earthquake engineering. Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver.
53. Moehle, J., & Deierlein, G. G. (2004, August). A framework methodology for performance-based earthquake engineering. In 13th world conference on earthquake engineering (Vol. 679, p. 12). WCEE Vancouver.
54. Nguyen, Q. T. (2022). Modal analysis-based damage detection and structural health monitoring approach for high-rise buildings, PhD thesis, Bursa Uludağ University Institute of Science. http://hdl.handle.net/11452/29150
55. Ni, Y. Q., Xia, H. W., Wong, K. Y., & Ko, J. M. (2012). In-service condition assessment of bridge deck using long-term monitoring data of strain response. Journal of Bridge Engineering, 17(6), 876-885.
56. Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures, 16(1), 73-98.

57. Peeters, B., & De Roeck, G. (2001). One-year monitoring of the Z24-Bridge: environmental effects versus damage events. Earthquake engineering & structural dynamics, 30(2), 149-171.
58. Plevris, V. (2024). Addressing the pitfalls of image-based structural health monitoring: a focus on false positives, false negatives, and base rate bias. arXiv preprint arXiv:2410.20384.
59. Riveros, C. A., García, E. F., & Rivero, J. E. (2014). Structural Damage Detection Using Sensors Optimally Located. World Academy of Science, Engineering and Technology, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering, 8(10), 1071-1074.
60. Rodríguez-Antuñano, I., Martínez-Sánchez, J., Lagüela-López, S., & Riveiro, B. (2022). Towards more resilient smart cities: Mt-insar monitoring of urban infrastructure using machine learning techniques. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, 10, 221-228.
61. Sevgen, Y. N., Kasımzade, A. A., Günday, F., Novruzov, M., Nematli, E., & Tunç, B. (2025, September). Structural Health Monitoring System and Providing Data to Control of Inteligent Building's Power Systems. In 2025 12th International Conference on Electrical and Electronics Engineering (ICEEE) (pp. 138-141). IEEE
62. Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. Mechanical systems and signal processing, 25(5), 1803-1836.
63. Sohn, H. (2007). Effects of environmental and operational variability on structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 539-560.
64. Sohn, H., Farrar, C. R., Hemez, F. M., et al. (2003). A review of structural health monitoring literature: 1996–2001. Los Alamos National Laboratory Report LA-13976-MS.
65. Soong, T. T., & Dargush, G. F. (1997). Passive energy dissipation systems in structural engineering. (No Title).
66. Spencer Jr, B. F., & Nagarajaiah, S. (2003). State of the art of structural control. Journal of structural engineering, 129(7), 845-856.
67. Spencer, B. F., & Nagarajaiah, S. (2003). State of the art of structural control. Journal of Structural Engineering, 129(7), 845–856.
68. Straub, D., & Der Kiureghian, A. (2010). Bayesian network enhanced with structural reliability methods: methodology. Journal of engineering mechanics, 136(10), 1248-1258.
69. Structurae. (2020, June 9). 400-ton Pendulum in the Onion. Structure. access date 14.02.2026 https://structurae.net/en/products-services/400-ton-pendulum-in-the-onion
70. Sun, X., Ilanko, S., Mochida, Y., & Tighe, R. C. (2023). A review on vibration-based damage detection methods for civil structures. Vibration, 6(4), 843-875.
71. Tellal, C. (2017). Statistical evaluation of displacement demands of reinforced concrete buildings obtained by time history analysis. Pamukkale University Institute of Science. https://hdl.handle.net/11499/2227
72. Wang, G., & Ke, J. (2024). Literature review on the structural health monitoring (SHM) of sustainable civil infrastructure: an analysis of influencing factors in the implementation. Buildings, 14(2), 402.
73. Wang, Y., Pei, Y., & Zhao, Y. (2008). Vibration-based damage detection with structural modal characteristics. The Baltic Journal of Road and Bridge Engineering, 3(1), 21-28.
74. Weber, F., Benicke, O., Huber, P., Distl, J., & Braun, C.(2019) Schwingungsreduktion von bauwerkwn: Schwinggungs- kompensation- isolation un- dampfung. 16. D-A-CH Tagung

Erdbebeningenieurwesen & Baudynamik (D-A-CH 2019) 26. und 27. September 2019 Universität Innsbruck

75. Weber, F., Huber, P., Spensberger, S., Distl, J., & Braun, C. (2019). Reduced-mass adaptive TMD for tall buildings damping. International journal of high-rise buildings, 8(2), 117-123.
76. Worden, K., & Manson, G. (2007). The application of machine learning to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 515-537.
77. Worden, K., Farrar, C. R., Manson, G., & Park, G. (2007). The fundamental axioms of structural health monitoring. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 463(2082), 1639-1664.
78. Yang, J., Peng, Y., Xie, J., & Wang, P. (2022). Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification. Sensors, 22(12), 4549.
79. Zacharakis, I., & Giagopoulos, D. (2022). Vibration-based damage detection using finite element modeling and the metaheuristic particle swarm optimization algorithm. Sensors, 22(14), 5079.
80. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things journal, 1(1), 22-32.
81. Zhang, J., Zeng, Y., & Starly, B. (2021). Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis. SN Applied Sciences, 3(4), 442.
82. Zhang, X., Liang, W., Yu, X., Xie, B., & Liu, C. (2026). Evaluating the impact of artificial intelligence development on urban resilience: Evidence from Chinese cities. iScience.
83. Zhang, Y., Zhang, C., Wang, S., Dui, H., & Chen, R. (2024). Health indicators for remaining useful life prediction of complex systems based on long short-term memory network and improved particle filter. Reliability Engineering & System Safety, 241, 109666.

# Synthesis of Terminal Synergetic Controller Based on Nonlinear State Observer for 2-DOF Helicopter System

Hoang Duc Long[1][0000-0003-4975-9547]

[1] Le Quy Don Technical University, Hanoi, Vietnam
longhd@lqdtu.edu.vn

**Abstract.** This paper presents the synthesis of a robust control scheme for a Two Degree-of-Freedom Helicopter System based on Terminal Synergetic Control combined with a Nonlinear State Observer. The 2DOFHS is a highly nonlinear, strongly coupled MIMO system characterized by parametric uncertainties and unknown external disturbances, which makes control design particularly challenging. To address these issues, a terminal synergetic control strategy is developed to ensure finite-time convergence of the pitch and yaw angles to their desired reference values, while guaranteeing closed-loop stability in the presence of bounded disturbances. Since the terminal synergetic control law requires full state information, a nonlinear state observer is designed to estimate the unmeasured state variables, namely the angular velocities of the pitch and yaw motions. The observer is constructed to achieve fast and accurate state estimation, enabling effective implementation of the proposed control law without direct measurement of all system states. Lyapunov-based stability analysis is carried out to rigorously prove finite-time convergence of the tracking errors and asymptotic convergence of the observer estimation errors. The effectiveness and robustness of the proposed controller–observer structure are validated through comprehensive simulation studies under different operating conditions, including set-point regulation and trajectory tracking scenarios, as well as in the presence of bounded external disturbances. The simulation results demonstrate fast transient response, high tracking accuracy, and strong disturbance rejection capability, confirming the superiority of the proposed approach for controlling nonlinear helicopter systems.

**Keywords:** 2-DOF Helicopter System, Terminal Synergetic Control, Nonlinear State Observer, Lyapunov function.

## 1 Introduction

The Two Degree-of-Freedom Helicopter System (2DOFHS), also known as the twin-rotor MIMO system, has been widely adopted as a benchmark platform for validating advanced control algorithms in both academic research and practical laboratory environments. This system exhibits several challenging characteristics that make control design nontrivial: (i) it is a highly nonlinear and strongly coupled multiple-input multiple-output (MIMO) system; (ii) it contains parametric uncertainties arising from

modeling simplifications and physical variations; (iii) it is subject to unknown external disturbances during operation; and (iv) it requires fast response and guaranteed stability around equilibrium points. Due to these properties, the 2DOFHS represents a realistic testbed for evaluating robustness, performance, and adaptability of modern control strategies.

In early studies, classical control approaches such as PID and linear quadratic regulator (LQR) control were applied to the linearized model of the 2DOFHS under relatively simple disturbance assumptions [1-4]. Although these methods are straightforward to implement and provide acceptable performance near operating points, they generally fail to account for the intrinsic nonlinear dynamics of the system and may suffer from degraded performance when subjected to large disturbances or parameter variations. Consequently, these linear control techniques are insufficient for achieving high-performance control over a wide operating range.

To overcome these limitations, various nonlinear and intelligent control methods have been proposed. Fuzzy logic control schemes were investigated in [5,6] to handle system nonlinearities and uncertainties. Sliding mode control (SMC) approaches were developed in [7-10] to enhance robustness against disturbances, while backstepping-based controllers were introduced in [11,12] to systematically stabilize the nonlinear dynamics. Furthermore, neural network-based control strategies were employed in [13,14] to approximate unknown nonlinearities. Although these methods significantly improve robustness and tracking performance, they often suffer from drawbacks such as chattering, high computational complexity, or sensitivity to tuning parameters.

In recent years, hybrid and adaptive control strategies have gained increasing attention in order to further enhance control performance. Adaptive backstepping control [15-17], adaptive neural control [18,19], and backstepping sliding mode control [20] have been proposed to address uncertainties and improve transient response. Despite their effectiveness, a common assumption in these approaches is the availability of full state measurements. In practical applications, however, not all state variables of the 2DOFHS-particularly angular velocities-are directly measurable, which necessitates the use of state observers.

Synergetic control theory, originally introduced by Kolesnikov [21], has emerged as an effective nonlinear control framework that ensures system stability through the design of invariant manifolds. Unlike sliding mode control, synergetic control eliminates chattering while retaining robustness properties. This control methodology has been successfully applied to a variety of nonlinear systems, including electromechanical systems and power converters [22-25]. More recently, Terminal Synergetic Control (TSC) has been developed as an extension of synergetic control, where terminal dynamics are employed to guarantee finite-time convergence of system states to desired equilibrium points or trajectories [26-29]. Finite-time convergence is particularly desirable in high-performance control applications, as it ensures fast response and improved tracking accuracy.

However, both synergetic and terminal synergetic control laws require accurate knowledge of the system states. To address this issue, several observer-based approaches for the 2DOFHS have been proposed, including sliding mode observers, nonlinear observers, and disturbance observers [30-33]. While these observers can

estimate unmeasured states, their integration with terminal synergetic control has not been sufficiently explored.

Motivated by the above observations, this paper proposes a novel control framework that combines Terminal Synergetic Control with a Nonlinear State Observer for the 2DOF Helicopter System in the presence of unknown bounded external disturbances. The nonlinear state observer is designed to estimate the unmeasured angular velocities, enabling full-state feedback implementation of the terminal synergetic controller. Lyapunov-based stability analysis is employed to rigorously prove finite-time convergence of the tracking errors and asymptotic convergence of the observer estimation errors. Comprehensive simulation studies are conducted to demonstrate the effectiveness, robustness, and fast convergence properties of the proposed approach under both set-point regulation and trajectory tracking scenarios.

The remainder of this paper is organized as follows. Section 2 presents the mathematical model of the 2DOF Helicopter System. Section 3 details the design of the terminal synergetic controller and the nonlinear state observer, along with stability proofs. Simulation results are discussed in Section 4 to validate the proposed method. Finally, conclusions and future research directions are provided in Section 5.

## 2 Mathematical model of 2-DOF Helicopter System

The Two Degree-of-Freedom Helicopter System consists of a rigid beam mounted on a pivot that allows rotational motion in both the vertical (pitch) and horizontal (yaw) planes. Two DC-motor-driven rotors are installed at the opposite ends of the beam: the main rotor generates lift to control the pitch motion, while the tail rotor produces lateral thrust to regulate the yaw motion, as illustrated in Figure 1.

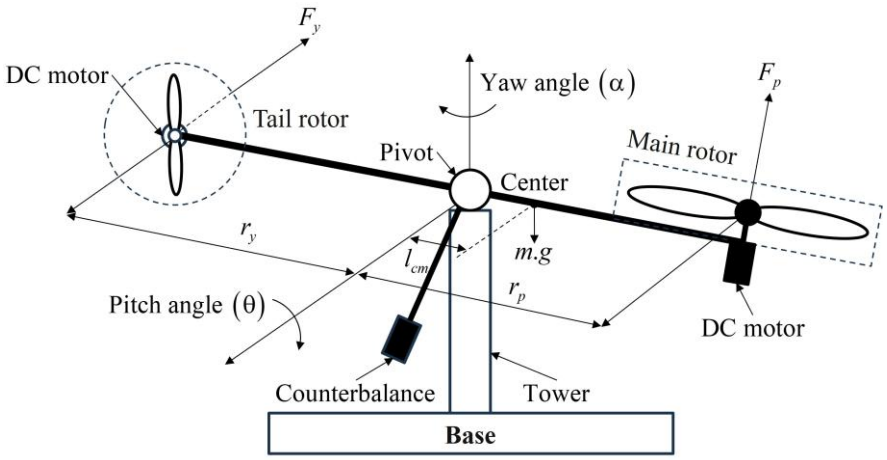



**Fig. 1.** 2-DOF Helicopter System.

A counterbalance mass is attached near the pivot to establish a stable equilibrium configuration and to compensate for gravitational effects. The entire assembly is fixed

on a tower structure, providing a safe and reliable experimental platform for control design and validation. Due to aerodynamic coupling between the rotors, gravitational forces, friction effects, and nonlinear inertial terms, the 2DOFHS exhibits strong nonlinear and cross-coupled dynamics. Let $\theta$ and $\alpha$ denote the pitch and yaw angles, respectively, while $V_p$ and $V_y$ represent the voltages applied to the main and tail DC motors.

The parameters and their values of TRMS are given in Table 1 [34].

**Table 1.** Parameters and values of TRMS.

| Parameter | Definition | Value |
|---|---|---|
| $\theta$ | The Pitch angle | - |
| $\alpha$ | The Yaw angle | - |
| $V_p$ | The voltage of main DC motor | - |
| $V_y$ | The voltage of tail DC motor | - |
| $J_p$ | Moment of inertia about the pitch axis | 0.0215 $(kgm^2)$ |
| $J_y$ | Moment of inertia about the yaw axis | 0.0237 $(kgm^2)$ |
| $l_{cm}$ | Distance between Center of mass and the body-fixed frame origin | 0.0071 $(m)$ |
| $m$ | Weight of the TRMS | 1.0750 $(kg)$ |
| $D_p$ | Pitch viscous friction constant | 0.0071 $\left(\frac{N}{V}\right)$ |
| $D_y$ | Yaw viscous friction constant | 0.0220 $\left(\frac{N}{V}\right)$ |
| $K_{pp}$ | Torque thrust gain acting on pitch axis from pitch propeller | 0.0011 $\left(\frac{Nm}{V}\right)$ |
| $K_{py}$ | Torque thrust gain acting on pitch axis from yaw propeller | 0.0021 $\left(\frac{Nm}{V}\right)$ |
| $K_{yp}$ | Torque thrust gain acting on yaw axis from pitch propeller | -0.0027 $\left(\frac{Nm}{V}\right)$ |
| $K_{yy}$ | Torque thrust gain acting on yaw axis from yaw propeller | 0.0022 $\left(\frac{Nm}{V}\right)$ |
| $g$ | Gravitational acceleration | 9.81 $\left(\frac{m}{s^2}\right)$ |

Based on Newton-Euler equations, the nonlinear dynamics of the 2DOFHS can be described as follows [8,16]:

$$\begin{aligned}\left(J_p + ml_{cm}^2\right)\ddot{\theta} &= K_{pp}V_p + K_{py}V_y - mgl_{cm}\cos\theta - D_p\dot{\theta} - ml_{cm}^2\dot{\alpha}^2\sin\theta\cos\theta \\ \left(J_y + ml_{cm}^2\cos^2\theta\right)\ddot{\alpha} &= K_{yp}V_p + K_{yy}V_y - D_y\dot{\alpha} + 2ml_{cm}^2\dot{\alpha}\dot{\theta}\sin\theta\cos\theta\end{aligned} \quad (1)$$

To simplify the notation and facilitate controller design, the following auxiliary variables are introduced:

$$\begin{aligned}a_1 &= J_p + ml_{cm}^2 \\ a_2 &= -mgl_{cm}\cos\theta - D_p\dot{\theta} - ml_{cm}^2\dot{\alpha}^2\sin\theta\cos\theta \\ b_1 &= J_y + ml_{cm}^2\cos^2\theta\end{aligned}$$

$$b_2 = -D_y\dot{\alpha} + 2ml_{cm}^2\dot{\alpha}\dot{\theta}\,sin\,\theta\,cos\,\theta \tag{2}$$

Using these definitions, the system dynamics can be rewritten in a compact form:

$$\begin{aligned}\ddot{\theta} &= \frac{a_2}{a_1} + \frac{K_{pp}V_p + K_{py}V_y}{a_1} \\ \ddot{\alpha} &= \frac{b_2}{b_1} + \frac{K_{yp}V_p + K_{yy}V_y}{b_1}\end{aligned} \tag{3}$$

To account for unknown external disturbances acting on the system, the state-space representation of the 2DOFHS is formulated as:

$$\begin{cases}\dot{x}_1 = f_1(x) + g_1(x)x_2 \\ \dot{x}_2 = f_2(x) + u_2(t) + d_2(t) \\ \dot{x}_3 = f_3(x) + g_3(x)x_4 \\ \dot{x}_4 = f_4(x) + u_4(t) + d_4(t)\end{cases} \tag{4}$$

where the state variables are defined as $x_1 = \theta$, $x_2 = \dot{\theta}$, $x_3 = \alpha$, $x_4 = \dot{\alpha}$. The system functions and control inputs are given by: $f_1(x) = f_3(x) = 0$, $g_1(x) = g_3(x) = 1$, $f_2(x) = \frac{a_2}{a_1}$, $u_2(t) = \frac{K_{pp}V_p + K_{py}V_y}{a_1}$, $f_4(x) = \frac{b_2}{b_1}$, $u_4(t) = \frac{K_{yp}V_p + K_{yy}V_y}{b_1}$.

The terms $d_2(t)$ and $d_4(t)$ represent unknown but bounded external disturbances satisfying $|d_i(t)| \le D_i$, $i = 2{,}4$. This state-space formulation highlights the nonlinear and coupled nature of the system and serves as the basis for the subsequent design of the terminal synergetic controller and nonlinear state observer.

## 3 Design of Terminal Synergetic Controller based on Nonlinear State Observer for 2-DOF Helicopter System

In this section, a terminal synergetic control scheme combined with a nonlinear state observer is developed for the 2-DOF Helicopter System subject to unknown bounded external disturbances. The proposed approach aims to achieve finite-time convergence of the tracking errors while guaranteeing closed-loop stability, even when not all state variables are directly measurable.

### 3.1 Design of the Terminal Synergetic Controller

According to the terminal synergetic control theory presented in [26-29], the controller design is based on the construction of macro variables that define invariant manifolds governing the closed-loop dynamics. These macro variables are selected such that the system trajectories are driven toward the desired equilibrium or reference trajectories in finite time.

For the position states, the first set of macro variables is defined as

$$\psi_i = x_i - x_{ir}, \quad i = 1{,}3 \tag{5}$$

where $x_{ir}$ denotes the desired reference trajectory of $x_i$.

For the velocity states, the second set of macro variables is defined as

$$\psi_i = x_i - \phi_i, \quad i = 2,4 \tag{6}$$

where $\phi_i$ represents internal virtual control inputs to be designed.

To enforce finite-time convergence, the terminal dynamic evolution of the macro variables is chosen as

$$\psi_i + T_i \dot{\psi}_i^{\frac{p_c}{q_c}} = 0 \tag{7}$$

where $T_i > 0$ for $i = \overline{1,4}$ are design parameters, $p_c$ and $q_c$ are positive odd integers satisfying $1 < \frac{p_c}{q_c} < 2$. This terminal dynamics guarantees that the macro variables converge to zero in finite time.

From (7), the macro-variable dynamics can be expressed as

$$\dot{\psi}_i = W_i \tag{8}$$

where $W_i = \left(-\frac{1}{T_i}\psi_i\right)^{\frac{q_c}{p_c}}$ for $i = \overline{1,4}$.

Substituting the system dynamics into the macro-variable definitions yields

$$\begin{cases} f_1(x) + g_1(x)x_2 - \dot{x}_{1r} = W_1 \\ f_3(x) + g_3(x)x_4 - \dot{x}_{3r} = W_3 \end{cases} \tag{9}$$

On the invariant manifold $\psi_i = 0$, it follows that $x_2 = \phi_2$ and $x_4 = \phi_4$. From (9)

$$\begin{cases} f_1(x) + g_1(x)\phi_2 - \dot{x}_{1r} = W_1 \\ f_3(x) + g_3(x)\phi_4 - \dot{x}_{3r} = W_3 \end{cases} \tag{10}$$

Hence, the internal control laws are obtained as

$$\begin{cases} \phi_2 = g_1^{-1}(x)(W_1 - f_1(x) + \dot{x}_{1r}) \\ \phi_4 = g_3^{-1}(x)(W_3 - f_3(x) + \dot{x}_{3r}) \end{cases} \tag{11}$$

Next, the external control inputs are designed to stabilize the velocity-related macro variables. From the system dynamics, one obtains

$$\begin{cases} f_2(x) + u_2(t) - \dot{\phi}_2 = W_2 \\ f_4(x) + u_4(t) - \dot{\phi}_4 = W_4 \end{cases} \tag{12}$$

Accordingly, the external control laws are given by

$$\begin{cases} u_2(t) = W_2 - f_2(x) + \dot{\phi}_2 + \varphi_2 - D_2\, sgn\, \psi_2 \\ u_4(t) = W_4 - f_4(x) + \dot{\phi}_4 + \varphi_4 - D_4\, sgn\, \psi_4 \end{cases} \tag{13}$$

where the additional coupling compensation terms are defined as $\varphi_2 = -g_1(x)\psi_1$ and $\varphi_4 = -g_3(x)\psi_3$.

**Theorem 1**. The nonlinear system (4) under the control laws (13) is locally stabilized, and the tracking errors converge to zero in finite time $T_s$, given by

$$T_s = t_0 + \frac{V^{1-\eta}(t_0)}{\beta(1-\eta)} \tag{14}$$

where $V = \frac{1}{2}\sum_{i=1}^{4}\psi_i^2$, $\eta = \frac{p_c+q_c}{2p_c}$, and $\beta = 2^{\eta}\min_{1\le i\le 4}\left(\frac{1}{T_i}\right)$.

**Proof**. Consider the Lyapunov candidate function

$$V = \frac{1}{2}\psi_1^2 + \frac{1}{2}\psi_2^2 + \frac{1}{2}\psi_3^2 + \frac{1}{2}\psi_4^2 \tag{15}$$

The derivative of the Lyapunov function is determined as

$$\dot{V} = \psi_1\dot{\psi}_1 + \psi_2\dot{\psi}_2 + \psi_3\dot{\psi}_3 + \psi_4\dot{\psi}_4 \tag{16}$$

Then

$$\begin{aligned}
\dot{V} &= \psi_1(\dot{x}_1 - \dot{x}_{1r}) + \psi_2\left(\dot{x}_2 - \dot{\phi}_2\right) + \psi_3(\dot{x}_3 - \dot{x}_{3r}) + \psi_4\left(\dot{x}_4 - \dot{\phi}_4\right)\\
&= \psi_1(f_1(x) + g_1(x)x_2 - \dot{x}_{1r}) + \psi_2\left(f_2(x) + u_2(t) + d_2(t) - \dot{\phi}_2\right)\\
&\quad + \psi_3(f_3(x) + g_3(x)x_4 - \dot{x}_{3r}) + \psi_4\left(f_4(x) + u_4(t) + d_4(t) - \dot{\phi}_4\right)\\
&= \psi_1(f_1(x) + g_1(x)(\psi_2 + \phi_2) - \dot{x}_{1r}) + \psi_2\left(f_2(x) + u_2(t) + d_2(t) - \dot{\phi}_2\right)\\
&\quad + \psi_3(f_3(x) + g_3(x)(\psi_4 + \phi_4) - \dot{x}_{3r}) + \psi_4\left(f_4(x) + u_4(t) + d_4(t) - \dot{\phi}_4\right)\\
&= \psi_1(f_1(x) + g_1(x)\psi_2 + g_1(x)\phi_2 - \dot{x}_{1r}) + \psi_2\left(f_2(x) + u_2(t) + d_2(t) - \dot{\phi}_2\right)\\
&\quad + \psi_3(f_3(x) + g_3(x)\psi_4 + g_3(x)\phi_4 - \dot{x}_{3r}) + \psi_4\left(f_4(x) + u_4(t) + d_4(t) - \dot{\phi}_4\right)
\end{aligned} \tag{17}$$

Substituting the internal controls (11) and external controls (13) into (17) yields

$$\begin{aligned}
\dot{V} &= \psi_1(f_1(x) + g_1(x)\phi_2 - \dot{x}_{1r} + g_1(x)\psi_2)\\
&\quad + \psi_2\left(f_2(x) + W_2 - f_2(x) + \dot{\phi}_2 + \varphi_2 - D_2\, sgn\,\psi_2 + d_2(t) - \dot{\phi}_2\right)\\
&\quad + \psi_3(f_3(x) + g_3(x)\phi_4 - \dot{x}_{3r} + g_3(x)\psi_4)\\
&\quad + \psi_4\left(f_4(x) + W_4 - f_4(x) + \dot{\phi}_4 + \varphi_4 - D_4\, sgn\,\psi_4 + d_4(t) - \dot{\phi}_4\right)\\
&= \psi_1(W_1 + g_1(x)\psi_2) + \psi_2\left(W_2 + \varphi_2 - D_2\, sgn\,\psi_2 + d_2(t)\right)\\
&\quad + \psi_3(W_3 + g_3(x)\psi_4) + \psi_4\left(W_4 + \varphi_4 - D_4\, sgn\,\psi_4 + d_4(t)\right)\\
&= \psi_1 W_1 + \psi_1 g_1(x)\psi_2 + \psi_2 W_2 + \psi_2\varphi_2 - \psi_2\left(D_2\, sgn\,\psi_2 - d_2(t)\right)\\
&\quad + \psi_3 W_3 + \psi_3 g_3(x)\psi_4 + \psi_4 W_4 + \psi_4\varphi_4 - \psi_4\left(D_4\, sgn\,\psi_4 - d_4(t)\right)
\end{aligned} \tag{18}$$

After substituting additional terms into (18) and note that $|d_i(t)| \le D_i$, $i = 2{,}4$:

$$\begin{aligned}
\dot{V} &= \psi_1 W_1 + \psi_1 g_1(x)\psi_2 + \psi_2 W_2 + \psi_2(-g_1(x)\psi_1) - \psi_2\left(D_2\, sgn\,\psi_2 - d_2(t)\right)\\
&\quad + \psi_3 W_3 + \psi_3 g_3(x)\psi_4 + \psi_4 W_4 + \psi_4(-g_3(x)\psi_3) - \psi_4\left(D_4\, sgn\,\psi_4 - d_4(t)\right)\\
&\le \psi_1 W_1 + \psi_2 W_2 + \psi_3 W_3 + \psi_4 W_4 - |\psi_2|\left(D_2 - d_2(t)\right) - |\psi_4|\left(D_4 - d_4(t)\right)\\
&\le \psi_1 W_1 + \psi_2 W_2 + \psi_3 W_3 + \psi_4 W_4
\end{aligned} \tag{19}$$

Because $p_c$ and $q_c$ are positive odd integers. Therefore,

$$\begin{aligned}\dot{V} &\le \psi_1\left(-\frac{1}{T_1}\psi_1\right)^{\frac{q_c}{p_c}} + \psi_2\left(-\frac{1}{T_2}\psi_2\right)^{\frac{q_c}{p_c}} + \psi_3\left(-\frac{1}{T_3}\psi_3\right)^{\frac{q_c}{p_c}} + \psi_4\left(-\frac{1}{T_4}\psi_4\right)^{\frac{q_c}{p_c}} \\ &= (-1)^{\frac{q_c}{p_c}}\left(\frac{1}{T_1}\right)^{\frac{q_c}{p_c}}(\psi_1)^{\frac{q_c}{p_c}+1} + (-1)^{\frac{q_c}{p_c}}\left(\frac{1}{T_2}\right)^{\frac{q_c}{p_c}}(\psi_2)^{\frac{q_c}{p_c}+1} \\ &+ (-1)^{\frac{q_c}{p_c}}\left(\frac{1}{T_3}\right)^{\frac{q_c}{p_c}}(\psi_3)^{\frac{q_c}{p_c}+1} + (-1)^{\frac{q_c}{p_c}}\left(\frac{1}{T_4}\right)^{\frac{q_c}{p_c}}(\psi_4)^{\frac{q_c}{p_c}+1} \\ &= -\left(\frac{1}{T_1}\right)^{\frac{q_c}{p_c}}(\psi_1)^{\frac{q_c}{p_c}+1} - \left(\frac{1}{T_2}\right)^{\frac{q_c}{p_c}}(\psi_2)^{\frac{q_c}{p_c}+1} \\ &\qquad - \left(\frac{1}{T_3}\right)^{\frac{q_c}{p_c}}(\psi_3)^{\frac{q_c}{p_c}+1} - \left(\frac{1}{T_4}\right)^{\frac{q_c}{p_c}}(\psi_4)^{\frac{q_c}{p_c}+1} \end{aligned} \quad (20)$$

Let $\lambda_i = \left(\frac{1}{T_i}\right)^{\frac{q_c}{p_c}}$, $i = \overline{1,4}$. We have

$$\begin{aligned}\dot{V} &\le -min\{\lambda_1,\lambda_2,\lambda_3,\lambda_4\}\left((\psi_1)^{\frac{q_c}{p_c}+1} + (\psi_2)^{\frac{q_c}{p_c}+1} + (\psi_3)^{\frac{q_c}{p_c}+1} + (\psi_4)^{\frac{q_c}{p_c}+1}\right) \\ &= -\lambda\left(2^{\frac{q_c+p_c}{2p_c}}\right)\left(\left(\frac{\psi_1^2}{2}\right)^{\frac{q_c+p_c}{2p_c}} + \left(\frac{\psi_2^2}{2}\right)^{\frac{q_c+p_c}{2p_c}} + \left(\frac{\psi_3^2}{2}\right)^{\frac{q_c+p_c}{2p_c}} + \left(\frac{\psi_4^2}{2}\right)^{\frac{q_c+p_c}{2p_c}}\right)_{min} \\ &= -\beta\left(\left(\frac{\psi_1^2}{2}\right)^{\eta} + \left(\frac{\psi_2^2}{2}\right)^{\eta} + \left(\frac{\psi_3^2}{2}\right)^{\eta} + \left(\frac{\psi_4^2}{2}\right)^{\eta}\right) \end{aligned} \quad (21)$$

where $\lambda\, min\{\lambda_1,\lambda_2,\lambda_3,\lambda_4\}_{min}$, $\beta = \lambda\left(2^{\frac{q_c+p_c}{2p_c}}\right)_{min}$, $\eta = \frac{q_c+p_c}{2p_c}$.

So that,

$$\dot{V} \le -\beta\left(\frac{\psi_1^2}{2} + \frac{\psi_2^2}{2} + \frac{\psi_3^2}{2} + \frac{\psi_4^2}{2}\right)^{\eta} = -\beta V^{\eta} \quad (22)$$

The inequality (22) implies that the settling time $T_s$ is determined as (14) [29].

From the definition of $u_2(t)$ and $u_4(t)$, the supplied voltages of DC motors are found

$$\begin{cases} V_p = \dfrac{(J_y + ml_{cm}^2 cos^2 x_1)u_4(t) - \frac{K_{yy}}{K_{py}}(J_p + ml_{cm}^2)u_2(t)}{K_{yp} - \frac{K_{yy}K_{pp}}{K_{py}}} \\ V_y = \dfrac{(J_p + ml_{cm}^2)u_2(t) - K_{pp}V_p}{K_{py}} \end{cases} \quad (23)$$

■

### 3.2 Design of Nonlinear State Observer

In practical implementations, the angular velocities of the pitch and yaw motions are often not directly measurable. Therefore, a nonlinear state observer is designed to estimate these unmeasured states.

By grouping the system states, the dynamics can be rewritten as

$$\begin{cases} \dot{x}_{13} = x_{24}, \\ \dot{x}_{13} = f(x_{24}) + g(x_{13})u(t) + d(t) \end{cases} \tag{24}$$

where $x_{13} = \begin{bmatrix} x_1 \\ x_3 \end{bmatrix}$, $x_{24} = \begin{bmatrix} x_2 \\ x_4 \end{bmatrix}$, and $d = \begin{bmatrix} d_2(t) \\ d_4(t) \end{bmatrix}$.

Let $\hat{x}_{13} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_3 \end{bmatrix}$ and $\hat{x}_{24} = \begin{bmatrix} \hat{x}_2 \\ \hat{x}_4 \end{bmatrix}$ denote the estimated state vectors of $x_{13}$ and $x_{24}$, respectively. The proposed nonlinear state observer is formulated as

$$\begin{cases} \dot{\hat{x}}_{13} = \hat{x}_{24} + K\text{sgn}(x_{13} - \hat{x}_{13}) \\ \dot{\hat{x}}_{24} \ = f(\hat{x}_{24}) + g(\hat{x}_{13})u + LK\text{sgn}(x_{13} - \hat{x}_{13}) \end{cases} \tag{25}$$

where $u = \begin{bmatrix} V_p \\ V_y \end{bmatrix}$, $K = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix}$ and $L = \begin{bmatrix} l_1 & 0 \\ 0 & l_2 \end{bmatrix}$ are positive definite observer gain matrices.

Assume $\tilde{x}_{13} = x_{13} - \hat{x}_{13}$ and $\tilde{x}_{24} = x_{24} - \hat{x}_{24}$. Then, the estimated error dynamics are expressed as

$$\begin{cases} \dot{\tilde{x}}_{13} = \tilde{x}_{24} - K\ sgn(\tilde{x}_{13}) \\ \dot{\tilde{x}}_{24} = f(x_{24}) - f(\hat{x}_{24}) + \big(g(x_{13}) - g(\hat{x}_{13})\big)u - LK\ sgn(\tilde{x}_{13}) \end{cases} \tag{26}$$

Define $c_2 = \frac{D_p}{a_1}, c_3 = \frac{ml_{cm}^2}{a_1}, c_4 = \frac{D_y}{b_1}$, and $c_5 = \frac{2ml_{cm}^2}{b_1}$, and assume that $c_2 \approx c_3 \approx c_4 \approx c_5 = c$ and $a_1 \approx b_1 = a$.

**Proposition 1**. Consider the observer (25) and assume that positive gains $\kappa_i$, $l_i$ for $i = 1,2$ satisfy

$$\begin{cases} \kappa_1 > \left|\dot{\tilde{\theta}}\right|, \kappa_2 > \left|\dot{\tilde{\alpha}}\right| \\ \left(l_1 + \frac{c}{a}\right)\left(l_2 + \frac{c}{a} + \zeta_1\right) > \zeta_2 \end{cases} \tag{27}$$

where $\zeta_1 = \frac{c}{a}\dot{\hat{\theta}}\ cos\ \theta\ sin\ \theta$ and $z_2 = \frac{c^2}{4a^2}\dot{\tilde{a}}^2 \cos^2 q \sin^2 q$. Then, the estimated error converges to zero.

**Proof**. Consider the observer error dynamics (26). During $\tilde{x}_1 = 0$ and thus $\tilde{x}_2 = K\ \widehat{sgn}\ x_1$. So that

$$\dot{\tilde{x}}_2 = -L\tilde{x}_2 - \begin{bmatrix} \frac{c}{a} & 0 \\ 0 & \frac{c}{a} \end{bmatrix}\tilde{x}_2 + cos\ \theta\ sin\ \theta \begin{bmatrix} -\frac{c}{a}\left(\dot{\alpha}^2 - \dot{\hat{\alpha}}^2\right) \\ \frac{c}{a}\left(\dot{\theta}\dot{\alpha} - \dot{\hat{\theta}}\dot{\hat{\alpha}}\right) \end{bmatrix} + \Delta_g u \tag{28}$$

where $\Delta_g = g(x_1) - g(\hat{x}_1) \approx 0$ because $\Delta_g$ only depends on $x_1$.

The Lyapunov function is chosen as

$$V(\tilde{x}_2) = \frac{1}{2}\left(\dot{\tilde{\theta}}^2 + \dot{\tilde{\alpha}}^2\right) \tag{29}$$

From (28) and (29)

$$\dot{V}(\tilde{x}_2) = -\left(l_1 + \frac{c}{a}\right)\dot{\tilde{\theta}}^2 - \left(l_2 + \frac{c}{a}\right)\dot{\tilde{\alpha}}^2 + \frac{c}{a}\cos\theta\sin\theta\left(-\dot{\tilde{\alpha}}\dot{\tilde{\theta}}\dot{\hat{\alpha}} + \dot{\tilde{\alpha}}^2\dot{\hat{\theta}}\right) \tag{30}$$

And after that

$$\dot{V}(\tilde{x}_2) = -\begin{bmatrix}\dot{\tilde{\theta}}\\ \dot{\tilde{\alpha}}\end{bmatrix}^T \begin{bmatrix} l_1 + \frac{c}{a} & -\frac{c}{2a}\dot{\tilde{\alpha}}\cos\theta\sin\theta \\ -\frac{c}{2a}\dot{\tilde{\alpha}}\cos\theta\sin\theta & l_2 + \frac{c}{a} + \frac{c}{a}\dot{\hat{\theta}}\cos\theta\sin\theta \end{bmatrix}\begin{bmatrix}\dot{\tilde{\theta}}\\ \dot{\tilde{\alpha}}\end{bmatrix} = -\dot{\tilde{x}}_2^T M \dot{\tilde{x}}_2 \tag{31}$$

where $\dot{\tilde{x}}_2 = \begin{bmatrix}\dot{\tilde{\theta}}\\ \dot{\tilde{\alpha}}\end{bmatrix}$ and $M = \begin{bmatrix} l_1 + \frac{c}{a} & -\frac{c}{2a}\dot{\tilde{\alpha}}\cos\theta\sin\theta \\ -\frac{c}{2a}\dot{\tilde{\alpha}}\cos\theta\sin\theta & l_2 + \frac{c}{a} + \frac{c}{a}\dot{\hat{\theta}}\cos\theta\sin\theta \end{bmatrix}$.

The conditions (27) implies that all the principal minors of $M$ are positive definite, and therefore $\dot{V}(\tilde{x}_2)$ is negative definite. Thus, the estimation error converges to zero at an exponential rate. Hence, the states $x_2$ are obtained exponentially fast. The observer gains $l_i (i = 1,2)$ are chosen relatively large so that the observer converges much faster than the system dynamics.

■

Combination of Terminal Synergetic Controller and Nonlinear State Observer, the control laws (23) can be written

$$\begin{cases} V_p = \dfrac{(J_y + ml_{cm}^2\cos^2(\hat{x}_1))u_4(t) - \frac{K_{yy}}{K_{py}}(J_p + ml_{cm}^2)u_2(t)}{K_{yp} - \frac{K_{yy}K_{pp}}{K_{py}}} \\ V_y = \dfrac{(J_p + ml_{cm}^2)u_2(t) - K_{pp}V_p}{K_{py}} \end{cases} \tag{32}$$

where the external control laws are $u_2(t) = W_2 - f_2(\hat{x}) + \dot{\phi}_2 + \varphi_2 - D_2\, sgn\,\psi_2$ and $u_4(t) = W_4 - f_4(\hat{x}) + \dot{\phi}_4 + \varphi_4 - D_4\, sgn\,\psi_4$, the internal control laws are $\phi_2 = g_1^{-1}(\hat{x})(W_1 - f_1(\hat{x}) + \dot{x}_{1r})$ and $\phi_4 = g_3^{-1}(\hat{x})(W_3 - f_3(\hat{x}) + \dot{x}_{3r})$, the additional terms are $\varphi_2 = -g_1(\hat{x})\psi_1$, $\varphi_4 = -g_3(\hat{x})\psi_3$, $D_i = max(|d_i(t)|)$ for $i = 2,4$.

In practical implementation and simulations, the discontinuous sign function $\mathrm{sgn}(\cdot)$ in nonlinear state observers (25), and the synergetic control law (32) is replaced by the function $\mathrm{sat}(\cdot)$ to alleviate chattering without affecting stability.

## 4 Simulation Results

In all simulation, the controller parameters are selected as $T_1 = T_2 = T_3 = T_4 = 0.08$, $p_c = 17$, $q_c = 15$ and and the observer gains are chosen as $\kappa_1 = \kappa_2 = 2$, $l_1 = l_2 = 40$ while the plant parameters follow Table 1. To assess robustness, the system is

excited by bounded random disturbances acting on the acceleration channels $d_2(t)$ and $d_4(t)$, with bounds $D_2 = D_4 = 2$. The initial values of state observer are $\left[\hat{\theta}(0) \quad \dot{\hat{\theta}}(0) \quad \hat{\alpha}(0) \quad \dot{\hat{\alpha}}(0)\right]^T$=$[0 \quad 0 \quad 0 \quad 0]^T$.

To show the effectiveness of the proposed controller (32), the author considered two cases:

**Case 1: Set-point regulation (constant references)**:

In the first scenario, the desired equilibrium is selected as $[\theta_d(t), \alpha_d(t)]^T = [1, 2]^T$ and the system starts from a displaced condition $\left[\theta(0), \dot{\theta}(0), \alpha(0), \dot{\alpha}(0)\right]^T = [2, 2, 3, 3]^T$.

Figures 2, 3 demonstrate that both outputs converge rapidly to their desired set points despite the presence of unknown disturbances, indicating strong disturbance rejection and effective decoupling behavior in the closed loop. Importantly, Figures 4, 5 show that the observer provides fast and accurate estimates: the estimation errors of both angles and angular velocities converge to (near) zero after a short transient, which validates the suitability of the NOS for implementing the full-state TSC law.

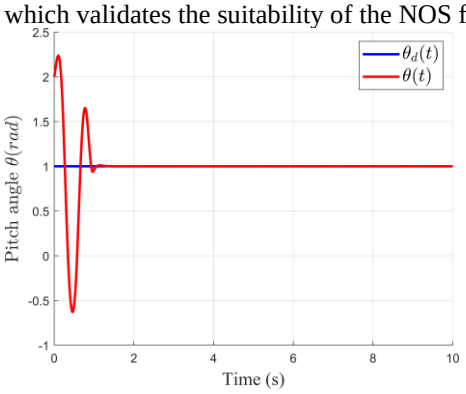



**Fig. 2.** The Pitch angle of Case 1.

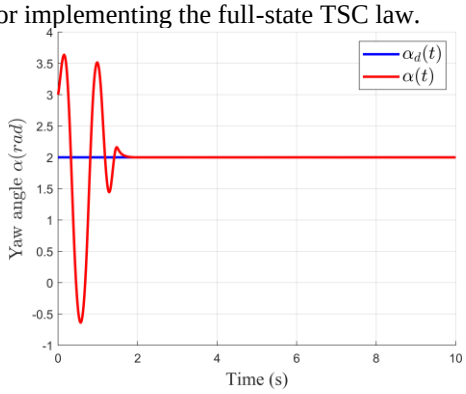



**Fig. 3.** The Yaw angle of Case 1.

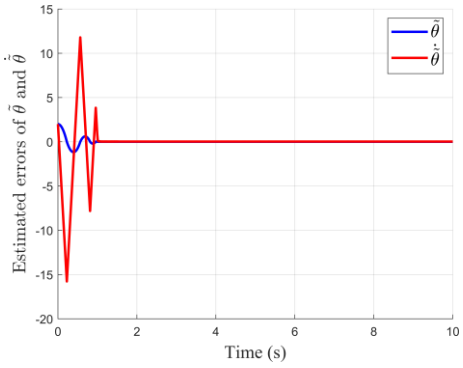



**Fig. 4.** Estimated errors of Pitch angle and its velocity of Case 1.

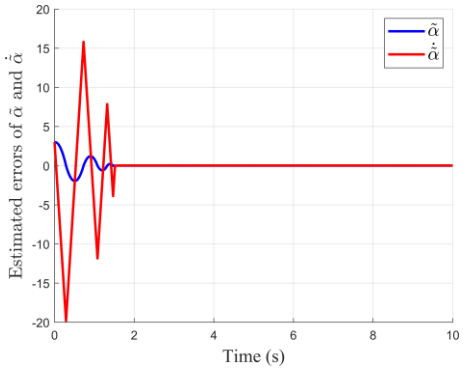



**Fig. 5.** Estimated errors of Yaw angle and its velocity of Case 1.

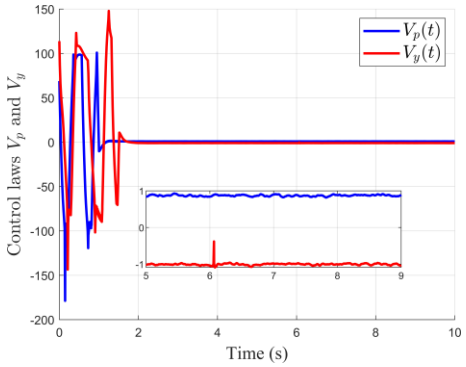



**Fig. 6.** The control laws of Case 1.

The corresponding control signals (Figure 6) remain bounded and exhibit a short initial transient needed to compensate for the large initial mismatch and disturbances; subsequently, the control effort decreases as the system approaches the invariant manifold and the tracking errors vanish.

Overall, the regulation results confirm that the proposed TSC-NOS structure achieves fast stabilization, high steady-state accuracy, and robustness against bounded perturbations.

**Case 2: Trajectory tracking (time-varying references)**:

To further evaluate tracking capability under time-varying commands, the references are chosen as smooth periodic trajectories: $[\theta_d(t), \alpha_d(t)]^T = [sin\, t\,, cos\, t]^T$ with the same initial state as in Case 1.

Figures 7, 8 show that $\theta(t)$ and $\alpha(t)$ accurately follow the desired trajectories with small tracking error, even during intervals where the reference curvature changes. The estimation-error plots (Figures 9, 10) verify that the observer continues to provide reliable velocity reconstruction in the tracking regime; after the initial transient, the estimated errors remain close to zero, which is critical for maintaining the intended finite-time convergence behavior of the terminal dynamics. The corresponding control inputs (Figure 11) are smooth and bounded, reflecting the inherent advantage of synergetic designs in avoiding the high-frequency switching typically observed in conventional sliding-mode implementations while still preserving robustness.

**Discussion**: Across both cases, the simulations collectively support the main theoretical claims: (i) the closed loop achieves rapid convergence of tracking errors, (ii) the observer errors decay quickly, enabling practical output-feedback implementation, and (iii) robustness is maintained under bounded disturbances using the compensation terms embedded in the controller design. The presented plots (outputs, estimation errors, and control voltages) therefore provide consistent evidence that the proposed TSC-NOS approach is an effective solution for the highly nonlinear, coupled 2-DOF helicopter dynamics under uncertainty.

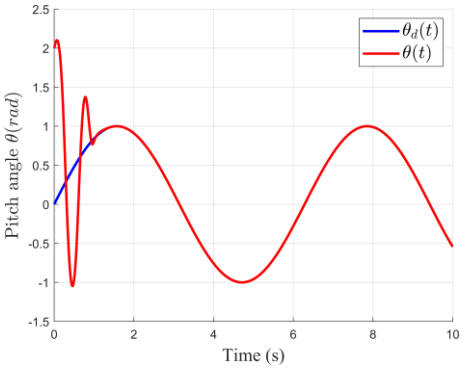



**Fig. 7.** Pitch angle of Case 2.

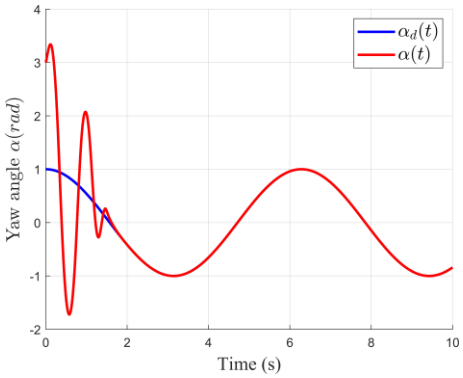



**Fig. 8.** Yaw angle of Case 2.

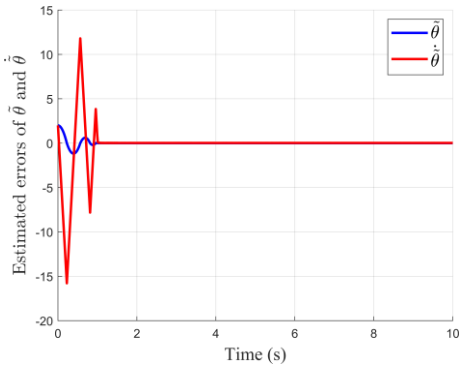



**Fig. 9.** Estimated errors of Pitch angle and its velocity of Case 2.

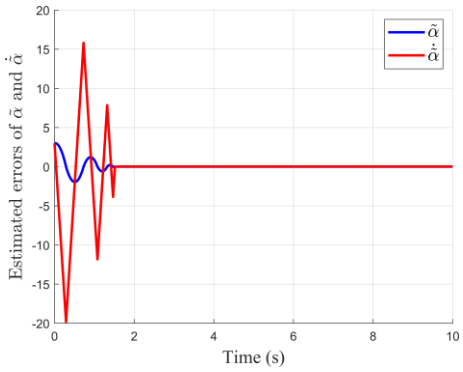



**Fig. 10.** Estimated errors of Yaw angle and its velocity of Case 2.

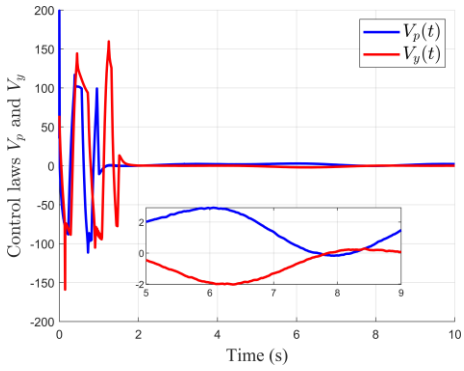



**Fig. 11.** The control laws of Case 2.

## 5 Conclusion

This paper has presented a robust observer-based control strategy for the Two Degree-of-Freedom Helicopter System by integrating Terminal Synergetic Control (TSC) with a Nonlinear State Observer. The considered system is characterized by strong nonlinearities, coupling effects, parametric uncertainties, and unknown bounded external disturbances, which pose significant challenges to conventional control techniques.

A terminal synergetic controller was systematically synthesized to ensure finite-time convergence of the pitch and yaw tracking errors to zero. The controller design was carried out using macro-variable construction and terminal dynamics, and its finite-time stability was rigorously established through Lyapunov-based analysis. To address the practical limitation of unavailable velocity measurements, a nonlinear state observer was designed to accurately estimate the angular velocities of the pitch and yaw motions. The convergence of the observer estimation errors was also analytically proven, guaranteeing reliable state reconstruction for output-feedback implementation.

Comprehensive simulation studies were conducted to validate the effectiveness and robustness of the proposed TSC-NOS framework. Both set-point regulation and trajectory-tracking scenarios were examined under unknown bounded disturbances. The results demonstrated fast transient response, high tracking accuracy, and strong disturbance rejection capability. Moreover, the observer exhibited rapid convergence, allowing the controller to maintain the desired finite-time performance without direct access to all state variables. The bounded and smooth control signals further confirm the practical feasibility of the proposed approach.

Overall, the presented results confirm that the proposed terminal synergetic control combined with a nonlinear state observer provides an effective and robust solution for controlling nonlinear helicopter systems with limited state measurements. Compared with existing approaches, the proposed method achieves improved convergence speed and robustness while avoiding excessive control effort.

Future research will focus on extending the proposed framework in several directions. First, the synthesis of an integral terminal synergetic controller will be investigated to further enhance steady-state performance in the presence of persistent disturbances. Second, optimal and adaptive tuning strategies will be explored to automatically select controller and observer parameters, thereby improving performance and reducing manual tuning effort. Finally, experimental validation on a real-time 2-DOF helicopter platform and consideration of actuator saturation constraints will be pursued to further demonstrate the applicability of the proposed method in practical control systems.

## References

1. Juang, J. G., Huang, M. T., Liu, W. K.: PID Control Using Presearched Genetic Algorithms for a MIMO System. In: *IEEE Transactions on Systems, Man, and*

*Cybernetics, Part C* (*Applications and Reviews*), vol. 38, no. 5, pp. 716-727 (2008), DOI: 10.1109/TSMCC.2008.923890.
2. Chaudhary, S., Kumar, A.: Control of Twin Rotor MIMO System Using 1-Degree-of-Freedom PID, 2-Degree-of-Freedom PID and Fractional order PID Controller. In: *2019 3rd International conference on Electronics, Communication and Aerospace Technology* (*ICECA*), pp. 746-751 (2019), DOI: 10.1109/ICECA.2019.8821923.
3. Phillips, A., Sahin, F.: Optimal control of a twin rotor MIMO system using LQR with integral action. In: *2014 World Automation Congress* (*WAC*), pp. 114-119 (2014), DOI: 10.1109/WAC.2014.6935709.
4. Gopmandal, F., Ghosh, A.: LQR-based MIMO PID control of a 2-DOF helicopter system with uncertain cross-coupled gain. In: *IFAC-PapersOnLine*, vol. 55, no. 22, pp. 183-188 (2022), DOI: 10.1016/j.ifacol.2023.03.031.
5. Islam, B. U., Ahmed, N., Bhatti, D. L., Khan, S.: Controller design using fuzzy logic for a twin rotor MIMO system. In: *7th International Multi Topic Conference, 2003. INMIC 2003.*, pp. 264-268 (2003), DOI: 10.1109/INMIC.2003.1416722.
6. Kafi, M. R., Chaoui, H., Miah, S.: Twin-rotor MIMO system and its control using interval type-2 fuzzy logic. In: *2017 IEEE International Symposium on Robotics and Intelligent Sensors* (*IRIS*), pp. 330-335 (2017), DOI: 10.1109/IRIS.2017.8250143.
7. Pratap, B., Purwar, S.: Sliding mode state observer for 2-DOF twin rotor MIMO system. In: *2010 International Conference on Power, Control and Embedded Systems*, pp. 1-6 (2010), DOI: 10.1109/ICPCES.2010.5698618.
8. Butt, S. S., Aschemann, H.: Multi-Variable Integral Sliding Mode Control of a Two Degrees of Freedom Helicopter. In: *IFAC-PapersOnLine*, vol. 48, no. 1, pp. 802-807 (2015), DOI: 10.1016/j.ifacol.2015.05.129.
9. Chithra, R., Thomas, K.: Robust Optimal Sliding Mode Control of Twin Rotor MIMO System. In: *International Journal of Engineering Research & Technology* (*IJERT*), vol. 5, no. 9, pp. 111-115 (2016).
10. Palepogu, K. R., Mahapatra, S.: Design of sliding mode control with state varying gains for a Benchmark Twin Rotor MIMO System in Horizontal Motion. In: *European Journal of Control*, vol. 75 (2024), DOI: 10.1016/j.ejcon.2023.100909.
11. Haruna, A., Mohamed, Z., Efe, M.Ö., Basri, M. A. M.: Dual boundary conditional integral backstepping control of a twin rotor MIMO system. In: *Journal of the Franklin Institute*, vol. 354, no. 15, pp. 6831-6854 (2017), DOI: 10.1016/j.jfranklin.2017.08.050.
12. Haruna, A., Mohamed, Z., Efe, M.Ö., Basri, M. A. M.: Improved integral backstepping control of variable speed motion systems with application to a laboratory helicopter. In: *ISA Transactions*, vol. 97, pp. 1-13 (2020), DOI: 10.1016/j.isatra.2019.07.016.
13. Shaik, F. A., Purwar, S. A.: A Nonlinear State Observer Design for 2-DOF Twin Rotor System Using Neural Networks. In: *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pp. 15-19 (2009), DOI: 10.1109/ACT.2009.219.
14. Aras, A. C., Kaynak, O.: Trajectory tracking of a 2-DOF helicopter system using neuro-fuzzy system with parameterized conjunctors. In: *2014 IEEE/ASME Inter-*

*national Conference on Advanced Intelligent Mechatronics*, pp. 322-326 (2014), DOI: 10.1109/AIM.2014.6878099.
15. Sodhi, P., Kar, I.: Adaptive Backstepping Control for A Twin Rotor MIMO System. In: *IFAC Proceedings Volumes*, vol. 47, no. 1, pp. 740-747 (2014), DOI: 10.3182/20140313-3-IN-3024.00082.
16. Patel, R., Deb, D., Modi, H., Shah, S.: Adaptive backstepping control scheme with integral action for quanser 2-dof helicopter. In: *2017 International Conference on Advances in Computing, Communications and Informatics* (*ICACCI*), pp. 571-577 (2017), DOI: 10.1109/ICACCI.2017.8125901.
17. Schlanbusch, S., Zhou, J.: Adaptive Backstepping Control of a 2-DOF Helicopter. In: *2019 7th International Conference on Control, Mechatronics and Automation* (*ICCMA*), pp. 210-215 (2019), DOI: 10.1109/ICCMA46720.2019.8988761.
18. Zou, T., Wu, H., Sun, W., Zhao, Z.: Adaptive neural network sliding mode control of a nonlinear two-degrees-of-freedom helicopter system. In: *Asian Journal of Control*, vol. 25, no. 3, pp. 2085-2094 (2023), DOI: 10.1002/asjc.2862.
19. Zhang, J., Yang, Y., Zhao, Z., Hong, K. S.: Adaptive Neural Network Control of a 2-DOF Helicopter System with Input Saturation. In: *International Journal of Control, Automation and Systems*, vol. 21, pp. 318-327 (2023), DOI: 10.1007/s12555-021-1011-2.
20. Wan, M., Chen, M., Lungu, M.: Integral Backstepping Sliding Mode Control for Unmanned Autonomous Helicopters Based on Neural Networks. In: *Drones*, vol. 7, no. 3, 154 (2023), DOI: 10.3390/drones7030154.
21. Kolesnikov, A. A.: Introduction of synergetic control. In: *2014 American Control Con-ference*, pp. 3013-3016 (2014), DOI: 10.1109/ACC.2014.6859397.
22. Asl, R. M., Hagh, Y. S., Anavatti, S., Handroos, H.: Adaptive finite integral nonsingular terminal synergetic control of nth-order nonlinear systems. In: *Mechanical Systems and Signal Processing*, vol. 142, 106789 (2020), DOI: 10.1016/j.ymssp.2020.106789.
23. Wang, K., Liu, D., Wang, L.: The Implementation of Synergetic Control for a DC-DC Buck-Boost Converter. In: *Procedia Computer Science*, vol. 199, pp. 900-907 (2022), DOI: 10.1016/j.procs.2022.01.113.
24. Saif, A., Fareh, R., Sinan, S., Bettayeb, M.: Fractional synergetic tracking control for robot manipulator. In: *Journal of Control and Decision*, vol. 11, no. 1, pp. 139-152 (2022), 10.1080/23307706.2022.2146008.
25. Long, H. D.: Design of Synergetic Controller Based on Nonlinear State Observer for Twin Rotor MIMO System. In: *Ronzhin, A., Truong, XT., Meshcheryakov, R.* (*eds*) *Interactive Collaborative Robotics. ICR 2025. Lecture Notes in Computer Science*(), vol. 16304, pp. 287-300 (2025), DOI: 10.1007/978-3-032-11903-2_22.
26. Jehar, D., Benmahammed, K., Cherif, A.: Terminal Synergetic Control of Crane System. In: *Journal of Engineering and Applied Sciences*, vol. 13, no. 3, pp. 3308-3313 (2018), DOI: 10.3923/jeasci.2018.3308.3313.
27. Rajendran, S., Diaz, M., Chavez, H., Cruchage, M., Castillo, E.: Terminal Synergetic Control for Variable Speed Wind Turbine Using a Two Mass Model. In: *2021 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies*, pp. 1-6 (2021), DOI: 10.1109/CHILECON54041.2021.9703058.

28. Rajendran, S., Jena, D., Diaz, M. and Rodrıguez, J.: Terminal Integral Synergetic Control for Wind Turbine at Region II Using a Two-Mass Model. In: *Processes*, vol. 11, no. 2, 616 (2023), DOI: 10.3390/pr11020616.
29. Babar. S. A., Rana, I. A., Mughal, I. S., Khan, S. A.: Terminal Synergetic and State Feedback Linearization Based Controllers for Artificial Pancreas in Type 1 Diabetic Patients. In *IEEE Access*, vol. 9, pp. 28012-28019 (2021), DOI: 10.1109/ACCESS.2021.3057365.
30. Saroj, D. K., Kar, I., Pandey, V. K.: Sliding mode controller design for Twin Rotor MIMO system with a nonlinear state observer. In: *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing* (*iMac4s*), pp. 668-673 (2013), DOI: 10.1109/iMac4s.2013.6526493.
31. Jadhav, T. S., Gadgune, S. Y.: Controller and Observer Techniques for Twin Rotor MIMO System. In: *International Journal of Engineering and Advanced Technology* (*IJEAT*), vol. 9, no. 1, pp. 595-590.
32. Ali, H., Ali, A., Shaikh, I. U.: Disturbance observer based control of twin rotor aerodynamic system. In: *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 4, pp. 2213-2227, DOI: 10.3906/elk-1912-34.
33. Ali, M. Z. H., Shaikh, I. ul H., Ali, A.: Disturbance Observer based Fractional Order Iterative Learning Control of Helicopter Model. In: *1st International Conference on Modern and Advanced Research*, pp. 353-363 (2023), DOI: 10.59287/icmar.1310.

# Digital Transformation of Rural Settlement Systems in Azerbaijan

Nagiyev Nizami[1][0000-0002-8160-6041], Aliyeva Sabina[2][0000-0002-0668-324X] and Aliyev Elbay[3][0009-0001-5734-2445]

[1] Azerbaijan University of Architecture and Construction, Head of the Department of "Fundamentals of Architecture," Honored Architect of Azerbaijan, Doctor of Architecture, Professor, Baku, Azerbaijan

[2] Azerbaijan University of Architecture and Construction, Lead Teacher at the Department of "Fundamentals of Architecture", Baku, Azerbaijan

[3] Azerbaijan University of Architecture and Construction, PhD (c), Assistant at the Department of "Fundamentals of Architecture", Baku, Azerbaijan
arc.elbay@gmail.com

**Abstract.** This article presents a comprehensive analysis of the paradigmatic transition in the planning, design and management of urban and rural settlements in the Republic of Azerbaijan. The study was conducted in the context of the implementation of the "Great Return" state program in the liberated territories and "Azerbaijan 2030: National Priorities for Socio-Economic Development". The work synthesizes the historical urbanization theories of "group settlement systems" described in detail in archival urban planning documents with modern achievements of Industry 4.0 - Building Information Modeling (BIM), Geographic Information Systems (GIS), Digital Twins and Cyber-Physical Systems (CPS). Special attention is paid to the analysis of the "Smart Village" pilot project in the Agali village of Zangilan region as a benchmark for post-conflict reconstruction. The article argues that the successful digital transformation of the Azerbaijani village requires not only the application of smart technologies, but also a deep integration of generative design with local (vernacular) architectural traditions. The research aims to develop a strategic roadmap for creating sustainable, culturally rooted, and economically viable rural communities. This approach can reverse the historical trend of "rural-urban" migration and ensure balanced territorial development.

**Keywords:** smart villages, smart cities, IT integrated architecture, Aghali village, Azerbaijan architecture

## 1 Introduction

In the age of global digital transformation, architectural and urban planning practices are undergoing profound changes. One of the key challenges in developing economies, including Azerbaijan, is the decline of rural settlements and the "rural-urban" imbalance against the backdrop of accelerating urbanization.[1] Rural communities, traditionally

based on an agrarian economy, are today in the process of being rebuilt with the help of information technology (IT), advanced digital services, and modern architectural approaches.

Historical settlement systems in Azerbaijan are not only a collection of residential areas, but also morphological structures that carry the socio-cultural code of the regions. However, during the period of industrialization and modernization, the mass influx towards the Baku and Absheron agglomerations led to demographic depletion in the regions and the marginalization of traditional rural infrastructure.[2] To prevent this process and restore the potential of villages, the Azerbaijani government has put forward the "Smart City" and "Smart Village" strategies within the framework of the "Great Return" state program, as well as in accordance with the document "Azerbaijan 2030: National Priorities for Socio-Economic Development".[3]

This study hypothesizes the role of the synthesis of IT and architecture in the revitalization of rural settlement systems in Azerbaijan. According to the hypothesis, improving the quality of life in villages and ensuring demographic sustainability is possible not only through the construction of physical buildings, but also through the complex management of space using Digital Twin, Internet of Things (IoT), and parametric architectural methods. The article compares global "Smart Village" trends with the specific geographical and cultural realities of Azerbaijan and evaluates the contributions of this approach to future urban planning science.

## 2 Problem Statement

The current situation of rural settlements in Azerbaijan consists of several fundamental problems of a consistent nature:

*Disruption of socio-spatial morphology and asymmetric urbanization.* Research by S. Naghiyev and I. Bayramov (2025) shows that the level of urbanization in Azerbaijan, which has reached 54.5%, is the result of continuous migration from rural to urban areas. The migration of young people and the working-age population to large cities for education and work purposes results in demographic aging and a decrease in the level of reproduction in villages.[2] The gap between economic and demographic urbanization indicators undermines the traditional archetypes of rural communities (production-domestic core).[1]

*Rural-urban service gap (Digital Divide).* The weakness of technological infrastructure (high-speed internet, automated utilities) and modern healthcare and education services in rural areas creates sharp inequalities in living standards.[5]

*Cultural and architectural degradation.* The spread of universal and standardized construction practices to villages threatens to lose local vernacular architectural traditions in regions with distinct climate and cultural zones, such as Sheki-Zagatala, Lankaran-Astara, and Absheron. Passive design strategies that ensure the adaptation of space to climate are being forgotten.[7 ]

*The need for complete reconstruction in post-conflict areas.* As a result of 30 years of occupation, thousands of settlements in the Karabakh and East Zangezur regions have been completely destroyed. Traditional restoration methods are not applicable in

these areas; there is a need for sustainable and innovative (Smart) solutions, based on the "tabula rasa" (scratchboard) principle.[9]

The research question is: How can information technologies and advanced architectural techniques be synthesized to ensure the economic independence of the village, preserve its traditional identity, and prevent migration flows?

## 3 The Main Part

This study is based on the "IT-integrated architecture" model, which proposes a conceptual framework for the digital transformation of rural settlement systems in Azerbaijan. The research methodology synthesizes traditional "socio-spatial morphology" theories with Industry 4.0 technologies. The main methodological approaches are as follows:

### 3.1 Spatial Planning with BIM and GIS Integration.

The proposed methodology is based on the scientific analysis of spatial data and is implemented at two main levels:

GIS Macro-Analytics. Geographic Information Systems (GIS) data are used for planning large areas during the restoration of Karabakh and the modernization of existing villages. These analyses allow to determine the topography, water resources, agricultural potential of the land and mine-hazardous zones. Also, based on this spatial analytics, the concept of "15-minute city/village" is applied.

BIM (Building Information Modeling) Micro-Analytics. BIM systems are applied for 3D modeling of engineering and architectural parameters of buildings.

### 3.2 Digital Twins.

As an extension of BIM, Digital Twin technology creates a virtual copy of both a single building and an entire village (for example, Agali). This tool is used to predict environmental problems, optimally manage resources (water, energy), and involve the population in participatory planning.

### 3.3 Parametric and Generative Design.

In order to preserve national-cultural identity, the modern concept of "Vernacular Digital Architecture" is applied with Generative Design and algorithmic approaches. This involves the integration of vernacular architectural traditions (e.g., Sheki grid and girikh ornaments) into modern buildings through parametric coding (e.g., through Grasshopper/Rhino).

Climate adaptation. The orientation of houses is optimized with aerodynamic simulations that provide appropriate internal courtyard systems, facade voids, and cross-

ventilation for the windy and sunny climates of Absheron and the humid climates of Lankaran.

Digital Fabrication. Grid ornaments are transformed into kinetic facades of buildings through Digital Fabrication and CNC cutting methods, which regulate the internal thermal comfort of buildings.

### 3.4 Cyber-Physical Systems (CPS) and Social Layer Application.

The sociological impact of the methodology is provided through Cyber-Physical Systems (CPS) and Community Hubs. CPS proactively manages the rural living environment by connecting physical buildings with a virtual database. Community Hubs, on the other hand, act as a Social Layer that prevents migration by facilitating young people's access to coworking, e-commerce, and remote work opportunities.

In modern urban planning, the morphological development of villages is no longer considered as isolated agrarian centers, but within the framework of "Group settlement systems". Analysis of archival and modern research1 shows that the resilience of villages depends on how they are integrated into larger agglomerations.

As noted by the Deputy Chairman of the State Committee for Urban Planning and Architecture, architect Ilgar Isbatov, Azerbaijan is a country distinguished by its historical urban planning and architectural traditions, and these traditions are being developed in the modern era through new institutional mechanisms and global platforms such as WUF13. According to historical dialectics, the village house should remain a "single production-household core". That is, digitalization should not eliminate traditional homestead farming, but rather strengthen it with the principles of "Smart Agriculture". This plays the role of a major buffer preventing migration.[1]

**Table 1.** Regional characteristics of architectural typology.

| Region | Climate characteristics | Traditional architectural response | Modern interpretation (Smart/Parametric) |
|---|---|---|---|
| **Lankaran-Astara** | Humid subtropical climate, abundant rainfall. | Sparse planning for cross ventilation. Orientation of streets along the latitude circle. Increasing the slope of roofs and water flows. | Aerodynamic modeling (CFD) of neighborhoods to maximize wind corridors. Pergola buildings with drainage systems integrated into the BIM model of the landscape. |
| **Absheron** | Semi-desert, strong northerly winds | Compact internal courtyards, blank north walls. Block | Parametric optimization of insolation. High-density cluster construction |

| | | | |
|---|---|---|---|
| | (Khazri), hot summer. | construction for wind protection. | with automated shading systems. Use of materials with high thermal mass. |
| **Mughan Plain** | Warm arid climate. | Exclusion of "mirror" arrangement of houses (window to window). Orientation of streets in the direction of the meridian. | Integration of rainwater collection systems into roofs. "Cool Roofs" and vertical greening to reduce the heat island effect. |
| **Sheki-Zaqatala** | Mild climate, rich craft traditions. | Use of river stone, brick, lattice stained glass. Large balconies. | Generative design of facades based on lattice patterns. Use of local materials as filler for 3D printing or modular blocks. |

### 3.5 "Smart Village" Paradigm: The Aghali Example.

The practical application of digital transformation of rural settlements in Azerbaijan is demonstrated in the village of Aghaly (Fig.1) in the Zangilan region (the "Smart Village" project). This project is built on 5 main pillars: smart infrastructure, smart governance, smart energy, smart education/healthcare, and smart agriculture.[11]
In Agali, buildings and streets are equipped with Internet of Things (IoT) sensors. Water, energy and road infrastructure are integrated into a "Smart Cloud Platform".[12] In agriculture, soil and vegetation are monitored using UAVs (drones) and multispectral cameras, and precision agriculture is applied, which dramatically reduces water and fertilizer consumption.[13]

All buildings in the project area are designed in accordance with the "zero emission" goal. Special thermoblocks were used for thermal insulation. The village's energy needs are met from alternative sources - Archimedes-type hydroturbines built on the Hakari River, solar panels (325 kW) and geothermal heat pumps.[15]

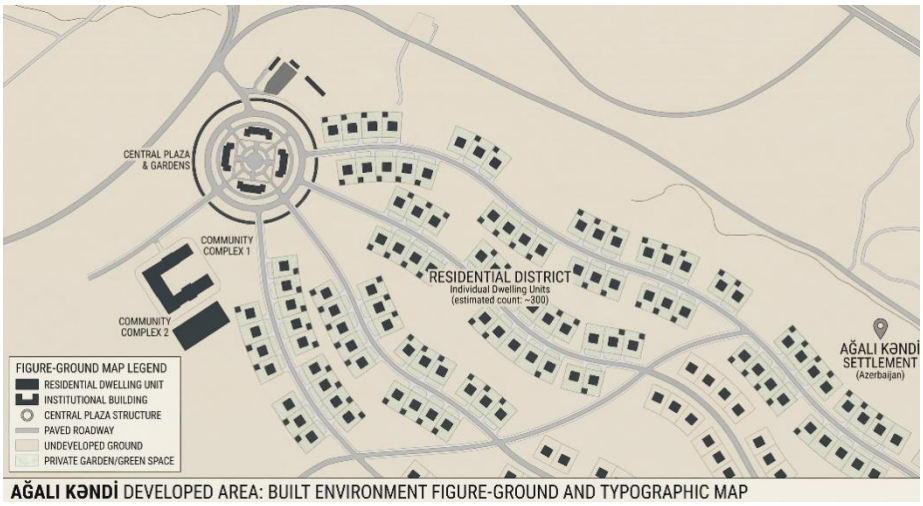



**Fig. 1.** Aghali Village. Built Environment figure-ground and typographic map

To increase the digital literacy of the community and facilitate access to public services, ASAN service kiosks and e-health (telemedicine) capabilities have been integrated into the village.[12] This connects rural residents to the global network, regardless of distance, and stimulates the activity of Community Hubs.

Spatial Planning with BIM and GIS Integration. The basis of digital transformation is the scientific analysis of spatial data. The "Improvised Methodology" proposed in the article is based on the integration of Geographic Information Systems (GIS) at the macro level and Building Information Modeling (BIM) at the micro level.[12]
During the reconstruction of Karabakh, as well as the modernization of existing villages, GIS data is being used to identify topography, water resources, agricultural potential of the land, and even mine-hazardous zones.[12] The implementation of the "15-minute city/village" concept is based on this spatial analytics.[14]

## 3.6 The Impact of Digital Transformation on Environmental and Sustainable Development

The digital transformation of rural settlement systems does not only serve demographic and economic purposes; it also forms the basis for environmental sustainability in Azerbaijan in order to achieve the Sustainable Development Goals (SDGs). The principles of "smart energy" and "green architecture", which are one of the main pillars of the Smart Village concept, are of critical importance in projects implemented in post-conflict areas (such as Agali). Technological tools not only reduce the energy demand of buildings through passive design strategies, but also fundamentally change the energy supply. Thanks to Internet of Things (IoT) sensors and Digital Twin platforms, the thermal efficiency, water consumption and electricity consumption of buildings are monitored in real time. This data is used, for example, to optimize cross-ventilation for the humid climate of Lankaran or to adjust facade gaps in the windy conditions of Absheron

(Parametric Design). The Agali example shows that by implementing alternative energy sources such as hydro turbines, solar panels and geothermal heat pumps on the Hakari River, rural settlements can become completely self-sufficient in terms of energy, "Net-Zero Energy" communities. This leads not only to economic savings, but also to a sharp reduction in carbon emissions, which is in line with Azerbaijan's Paris Agreement commitments.

Against the backdrop of urbanization, the revival of the rural economy is possible at the same time by reducing its ecological footprint. Precision Agriculture systems play a critical role, especially in areas that have been liberated from occupation, are at risk of landmines and need to restore soil quality. GIS macro-analytics allows for the precise identification of agriculturally suitable areas, water resources and micro-climatic zones. Monitoring of vegetation and soil health with multispectral cameras carried out by UAVs (drones) reduces irrigation and fertilizer consumption to the minimum needed. This is invaluable in terms of managing water resources (vital for Azerbaijan's arid climate) and limiting the use of chemicals. Thus, agriculture is transforming into a digital ecological farming model while preserving the tradition of "production-consumption core".

### 3.7 Community Hubs and the Shaping of the New Rural Economy

The lack of economic opportunities and weak social infrastructure, which are the main reasons for rural-urban migration, can be addressed through the concept of "Community Hubs" proposed through IT-integrated architecture. Acting as a functional continuation of the traditional village center in the digital age, these hubs are not just physical buildings, but also social and economic ecosystems managed by Cyber-Physical Systems (CPS).

Community Hubs, equipped with high-speed internet infrastructure, enable rural youth and the working population to find competitive jobs in the global market. These centers act as co-working spaces for remote workers, digital literacy courses, and startup incubation zones. This approach eliminates the need for young people to move to Baku or other large cities to find jobs, thus directly increasing demographic sustainability in villages.

The global marketing of agricultural and handicraft products through e-commerce platforms is a key element for the diversification of rural economies. Community Hubs assume a central function for the digital marketing, packaging and logistics management of these products. The integration of Cyber-Physical Systems ensures transparency and efficiency of the entire chain from production to sale, which increases trust in rural products.

As a result, this approach turns villages not just settlements, but also centers of the "Smart Spatial Economy". Strengthening social capital and community spirit through technological investments is seen as a key mechanism to reverse migration trends.

The engineering and architectural parameters of buildings are modeled in 3D using BIM. A step further, Digital Twins create a virtual copy of both a single building and an entire village (for example, Agali). This technology, which has been tested in China and European countries (for example, "Digital Twin of the Ocean" or village ecological

monitoring platforms), is an important tool in Azerbaijan for predicting environmental problems, optimal management of resources (water, energy), and involving the population in participatory planning.[14]

Preserving Local Traditions in Modern Parametric Design. Digitalization is not just about engineering solutions; it also ensures the transmission of national-cultural identity into the future through architecture. Different geographical zones of Azerbaijan, such as Sheki-Zagatala, Absheron and Lankaran, have their own vernacular (folk) architectural traditions adapted to their unique climate.[1]

As emphasized in the research conducted by Aydan Agabeyli (2017), the network (a geometric unity of colored glass and wood elements without glue and nails), which is the pearl of Sheki architecture, is based on complex mathematical and geometric calculations.[7] Within the modern concept of "Vernacular digital architecture", these traditions are reinterpreted through Generative Design and algorithmic approaches.

The inner courtyard system and specific facade spaces can be optimized against strong winds (khazri and gilavar) and high solar radiation with parametric coding (e.g. Grasshopper/Rhino).

With digital fabrication and CNC cutting techniques, traditional grid and girix ornaments can be transformed into kinetic facades of modern public buildings and village centers. These facades automatically open and close according to the angle of the sun's rays, regulating the internal thermal comfort of the buildings.

With aerodynamic simulations to ensure cross ventilation, the location and orientation architecture of local houses are adapted to modern energy efficiency standards (Class A).[8]

The Sociological Impact of Cyber-Physical Systems (CPS) and Community Hubs. The power of the proposed IT architecture model to reduce rural-urban migration depends in particular on the functioning of the Social Layer. Cyber-physical systems connect physical buildings with virtual databases, managing the rural living environment proactively rather than reactively.[7] The benefits of these systems to society are delivered through Community Hubs. The former traditional "teahouse" or village center concept is being transformed into digital "Community Hubs".[7] This hubs provides access to co-working, e-commerce (exporting local products to the global market), and remote work opportunities for the population, especially young people. As a result, young people do not have to move to Baku to find work; they can engage in full-fledged economic and social activities within the "Smart Village".[13]

The discussion of large-scale rural digitalization shows that technology is not an end in itself, but a tool for ensuring socio-economic sustainability. In the settlement morphology of Azerbaijan, this approach includes economic profitability (increasing agricultural production thanks to drones and IoT)[12] and improving the quality of social life. Gasimzadeh's (2023) Cost-Benefit Analysis also confirms that, although the initial costs of Agali-type projects are high, they are invaluable in terms of long-term development, energy savings, and social reintegration. However, the main threat to be addressed is the "digital divide". In addition to building advanced infrastructure, increasing the digital literacy of the rural population and investing in human capital are of vital importance.

**Table 2.** Comparative analysis of approaches to rural construction

| Parameter | Traditional Soviet planning | Modern "Smart Village" approach (Agaly) | Perspective approach (2030+) |
|---|---|---|---|
| Settlement structure | Monocentric, rigid hierarchy | Cluster, autonomous | Network, polycentric (Group Systems) |
| Energetics | Centralized (gas/electricity) | Hybrid (Green Energy + grid) | Distributed (Prosumers), micro-grids |
| Architecture | Typical, unified | Energy efficient, modern | Parametric, neo-vernacular |
| Governence | Bureaucratic, paper-based | Electronic (Smart Management) | Predictive (Digital Twin + artificial intelligence) |
| Economics | Collective Farm / Individual | Agroparks + Processing | Precision farming + E-commerce |

## 4 Conclusion

According to the results, the practical application of digital transformation of rural settlement systems in Azerbaijan is demonstrated by the "Smart Village" pilot project in Aghaly village of Zangilan, which is built on five main pillars: smart infrastructure, governance, energy, education/health, and agriculture. Technologically, buildings and streets are equipped with Internet of Things (IoT) sensors, and energy and road infrastructure are integrated into a smart cloud platform; in agriculture, precision farming is applied using UAVs and multispectral cameras, drastically reducing water and fertilizer consumption. At the same time, the project achieved high energy efficiency by designing all buildings in accordance with the zero emission target and providing energy demand from alternative sources - hydro turbines, solar panels, and geothermal heat pumps. Spatial planning is based on the integration of GIS analytics (resource, topography, and mine hazard zones) at the macro level and BIM/Digital Twins systems at the micro level, serving optimal resource management and prediction of environmental problems. In the field of architecture, local vernacular (folk) traditions, such as grid elements, were synthesized with Generative Design and parametric coding, creating

kinetic facades and orientation architecture that provide energy efficiency, suitable for the winds of Absheron or the humid climate of Lankaran. From a social perspective, the implementation of Cyber-Physical Systems such as ASAN service kiosks, e-health and digital Community Hubs not only increases the digital literacy of the community, but also plays an important buffer role in preventing rural-urban migration by providing young people with access to local coworking, e-commerce and remote work opportunities. Overall, the conclusion of the study is that this IT-integrated architectural model combines historical settlement forms with the modern "Smart Village" concept and is fully consistent with Azerbaijan's "Great Return" strategy.

## References


1. Aliyev, I. (2008). Rural-urban labour migration in Azerbaijan. Thesis Study.
2. Nagiyev, S. G., & Bayramov, I. A. (2025). Urban Settlement Patterns and Socio-Economic Development of Cities in the Republic of Azerbaijan. Journal of Earth Sciences & Environment, 2(3).
3. Guliyeva, G., & Museyibov, A. (2023). "SMART CITY AND SMART VILLAGE" CONCEPT IN KARABAKH REGION OF AZERBAIJAN: POLICY AND REALIZATION. International Conference on Economics.
4. İsbatov, İ. A. (2025). Azərbaycan şəhərsalma və memarlıq ənənələri. Dövlət Şəhərsalma və Arxitektura Komitəsinin hesabat materialları.
5. Fikrətzadə, F., & Hacıyeva, S. (2020). İşğaldan azad olunan ərazilərimizdə kənd təsərrüfatı sahəsinin bərpası istiqamətləri və istehsal göstəricilərinin proqnozlaşdırılması. Kənd təsərrüfatının iqtisadiyyatı, (4), 23-37.
6. Qasımzadə, A. (2023). Ağıllı şəhər layihə idarəetməsinin xərc-fayda təhlili: Ağalı ağıllı kənd layihəsi nümunəsi. UNEC Biznes Məktəbi.
7. Aghabayli, A. (2017). Geometric Patterns In Islamic decoration - A Parametric Envision of Portuguese and Azerbaijan Islamic Geometric Motifs. Faculty of Architecture, University of Lisbon.
8. Ebrahimi, A. N., & Rahimian, F. P. (2013). Impacts of climate on genesis of vernacular architecture of different parts of Iran: case study of cold and dry Azerbaijan. ALAM CIPTA, 6(1), 69-82.
9. Safarov, B. A. (2023). "Ağıllı kənd" - ölkəmiz üçün yeni təcrübə.
10. Ismayilova, N. (2022). The Smart Village Concept and Its Applicability in the Karabakh Region: Can It Succeed? Caucasus Analytical Digest, No. 125.
11. Guliyev, F. (2025). Post-conflict Resettlement in Karabakh: Rebuilding Livelihoods. Researching Internal Displacement.
12. Basir, W. N. F. W. A., et al. (2020). The Integration of BIM and GIS in Construction Project - A Data Consistency Review. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
13. Baghirova, N. (2023). Sustainable Agriculture in Aghali: The Smart Village Concept in the Great Return to Karabakh. Baku Dialogues.
14. Tan, F., & Cheng, Y. (2024). A digital twin framework for innovating rural ecological landscape control. Environmental Sciences Europe, Springer Nature, 36:59.
15. Azərbaycan Respublikası Prezidentinin Administrasiyası (2021). "Azərbaycan 2030: sosial-iqtisadi inkişafa dair Milli Prioritetlər".

# A Search and Rescue Robot for Under Rubble Victim Detection Using Thermal and Sensor Based Systems

Bensu Melis ORHAN[1] and İrem KAYAARSLAN[2]

[1,2] Baskent University, Electrical and Electronics Engineering, Ankara, Türkiye
[1]melisorhan063@gmail.com, [2]kyrslnirem@gmail.com

In the critical hours following a natural disaster, achieving a rapid situational assessment is required for the search and rescue operations carried out after natural disasters, especially in collapsed structures. The proposed robotic system provides simultaneous location determination and mapping (SLAM), it uses the robot operating system (ROS2) as a monitoring system for detecting victim signatures. Additionally, it facilitates durability with its robotic platform and track-type system.

Within our integrated design, we combined 2D LiDAR and an MLX90640 thermal sensor using a fusion strategy. This allows the robot to perform spatial mapping while simultaneously detecting survivors by analyzing temperature differentials in its surroundings.

Furthermore, a USB camera was integrated to provide the operator with a real time visual perspective. Data from all integrated sensors, alongside mechanical motion control, are centralized within the RViz2 framework, offering a unified operational dashboard for comprehensive visualization.

Communication is maintained via a high bandwidth TCP/IP network over a Wi-Fi module, ensuring seamless telemetry and real time sensor data streaming. For humanitarian support, the robot incorporates a walkie-talkie voice link alongside a servo mechanism designed to deliver essential life support supplies to trapped individuals. Experimental evaluations demonstrate that the system's performance in spatial reconstruction and thermal signature detection remains resilient against environmental noise. The integration of ROS2 and RVIZ2 ensures a scalable, modular, and professional framework for modern disaster response initiatives.

**Keywords:** Search and Rescue Robotics, ROS2, Sensor Fusion, SLAM, Thermal Imaging, Disaster Response.

# 1 Introduction

Our research is motivated by the critical need for resilient search and rescue (SAR) technologies in the immediate aftermath of seismic disasters. When buildings collapse, the resulting environment often becomes highly unstable and inaccessible for human responders, while the first “golden hours” remain crucial for locating and rescuing survivors. To mitigate these risks, this study proposes a tracked robotic platform designed to navigate high-friction debris fields where conventional wheeled or legged systems frequently experience mobility limitations. [1]

One of the primary challenges in such environments is the loss of reliable sensory perception due to dense dust, low visibility, and the absence of ambient light. Conventional RGB cameras therefore become ineffective in many SAR scenarios. To address this limitation, the proposed system employs a dual layer sensing architecture consisting of a 2D LiDAR sensor for real time Simultaneous Localization and Mapping (SLAM) [2] and an MLX90640 thermal sensor array for detecting human heat signatures [3]. By combining LiDAR-based environmental perception with thermal data analysis, the robot is capable of generating navigable maps while simultaneously identifying potential victims even in visually obscured environments. This integrated sensing approach enhances situational awareness and enables rescue teams to obtain critical information from hazardous areas without requiring direct human intervention.

## 2 Related Work

Search and rescue robotics has received increasing attention due to its potential to reduce the risks faced by human responders in hazardous post-disaster environments. Mobile robotic systems are particularly valuable in collapsed structures, where unstable debris, narrow passages, and limited accessibility make direct intervention difficult. In this context, Pinto et al. presented a low cost robotic system for search and rescue applications and demonstrated that practical robotic platforms can support disaster response tasks under constrained operational conditions. Their work highlights the importance of compact and affordable robotic solutions for field deployment in emergency scenarios [4].

Thermal imaging has also been widely studied as an effective method for victim detection in disaster environments. Since conventional RGB cameras often fail under smoke, darkness, or dense dust, infrared-based sensing provides a more reliable alternative for identifying human presence through body heat. Adams and Jose investigated the automated detection of trapped victims using thermal imaging and showed that thermal sensing can significantly improve victim localization in visually degraded conditions [5]. Similarly, the report cites the work of Pinto et al., where the MLX90640 sensor was used to detect human bodies based on temperature differences from the surrounding environment, confirming the suitability of thermal sensing for search and rescue robots [4].

In addition to victim detection, mapping and navigation remain essential capabilities for robotic rescue systems. LiDAR-based perception and SLAM methods are widely used for enabling robots to operate in unknown and cluttered environments. Bahraini et al. discussed mobile robot navigation and mapping in unknown environments using SLAM techniques, emphasizing the importance of simultaneous localization and mapping for safe movement and environmental awareness [6]. In parallel, Macenski et al. introduced SLAM Toolbox as a practical framework for real-time SLAM in dynamic robotic applications [7], while ROS2-based visualization and mapping tools such as RViz2 and slam toolbox provide effective infrastructure for implementing and monitoring these processes in practice [8], [9].

However, most existing studies address either victim detection or autonomous navigation separately. In contrast, the proposed system integrates LiDAR based SLAM and thermal sensing within a unified ROS2 framework to support both navigation and victim localization in disaster environments.

# 3 System Architecture

The core design of our platform focuses on keeping the robot functional even when disaster environments become unpredictable. We decided against a monolithic software structure, opting instead for a decoupled, modular framework. This specific architecture ensures that high bandwidth processes, like 2D spatial mapping, do not starve the primary locomotion control of resources. By separating these layers, we ensure the robot remains responsive to safety and driving commands, even when the main processor is under heavy computational load.

## 3.1 General Design of the Robot

We chose a tracked chassis primarily because post disaster zones are rarely uniform. Unlike wheels that often lose traction on loose debris or dust, tracks provide the surface area needed to climb obstacles while maintaining longitudinal stability. To improve the overall center of gravity, we housed the batteries and drive electronics within the lower section of the frame. A vertical mast was added to support the 2D LiDAR and the MLX90640 thermal array.

## 3.2 System Block Diagram

To prevent data bottlenecks, the internal processing network follows a distinct two layer hierarchy. The overall system architecture and the interaction between the hardware components are illustrated in Fig. 1.

The Raspberry Pi 4 runs on an Ubuntu Linux environment and acts as the coordinator for all data intensive operations within the ROS2 Humble framework. It handles the processing of LiDAR point clouds, executes SLAM algorithms for positioning, and interprets the thermal matrix via a Python based pipeline.

L298N Motor Driver our design deliberately avoids an extra microcontroller for driving the motors to prevent unnecessary system bloat. We instead wired the L298N H-Bridge driver straight into the Raspberry Pi's GPIO pins. This setup lets the motor driver handle those high current spikes from the DC motors while converting raw logic into physical torque.

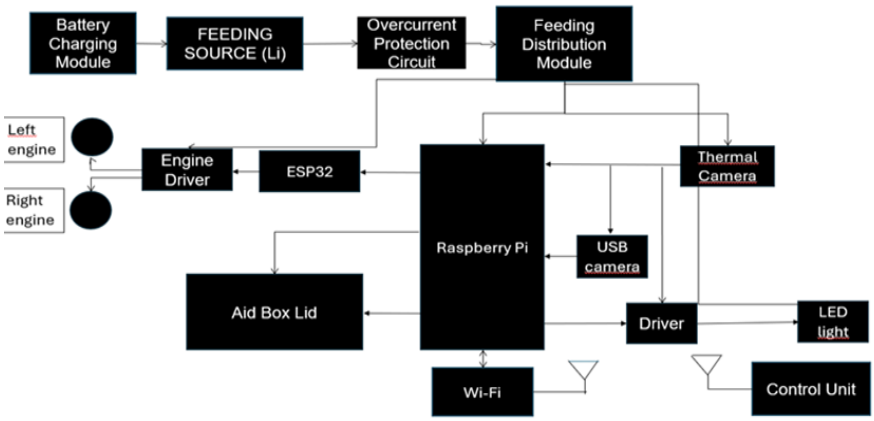



**Fig. 1.** The integrated hardware architecture of the robotic system.

## 4 Hardware Design

The robotic platform's physical design focuses on maintaining data integrity while navigating the unpredictable terrain of disaster zones. Instead of relying on a generic assembly, we developed a modular hardware stack, ensuring each peripheral is calibrated for the specific constraints of search and rescue operations.

The internal hardware components of the robotic platform and the Raspberry Pi processing unit used in the system are shown in Fig. 2. and Fig. 3.

### 4.1 Raspberry Pi Control Unit

A Raspberry Pi 4 serves as the primary compute module, managing high-level navigation while fusing multi modal sensor inputs. To meet the low-latency requirements of search and rescue operations, the unit coordinates real-time communication across all hardware interfaces. System wide synchronization is managed via the ROS 2 (Robot Operating System) framework, which integrates the LiDAR, thermal camera, and motor controllers into a unified data stream. This architecture enables concurrent environmental mapping and victim detection, providing the necessary throughput for complex autonomous tasks.

### 4.2 MLX90640 Thermal Camera

For survivor detection in low-visibility environments, the system utilizes an MLX90640 thermal sensor [10]. By measuring infrared radiation instead of visible light as conventional RGB cameras do, this unit converts object emissions into temperature based spatial data. In disaster zones where dust, smoke, or total darkness renders visual sensors ineffective, such thermal functionality becomes critical. Operating with a

32×24 pixel array, the sensor provides sufficient resolution to distinguish human body temperatures from the ambient environment. The resulting telemetry is processed via Python based algorithms for heatmap generation, facilitating the extraction of victim coordinates during active search missions.

### 4.3 LiDAR Integration and SLAM Mapping

Spatial awareness is derived from an RPLIDAR C1 unit, configured for a 360 degree environmental sweep [11]. This scanner feeds high density distance data directly into the SLAM pipeline. Our integration focused on converting raw reflections into 2D occupancy grids in real-time. By utilizing the C1's high sampling frequency, we successfully mapped jagged rubble and thin structural hazards.

### 4.4 Motor Driver and Mobility System

Navigating unpredictable disaster zones requires the DC motor-driven tracked system utilized by this platform. To manage motion, a driver module translates logic level signals from the Raspberry Pi into the high current electrical outputs necessary for actuation. This tracked configuration ensures superior traversal across uneven or debris laden terrain, offering a distinct mechanical advantage over conventional wheeled platforms. Given the constant variability inherent in disaster environments, such stability remains a fundamental prerequisite for consistent mobility and overall mission success.

### 4.5 Power Supply and Battery System

System power is supplied by a rechargeable lithium-ion battery pack, specifically rated to sustain the Raspberry Pi, sensing suite, and propulsion mechanism throughout rescue operations. A battery charging module is integrated into the system to allow convenient recharging of the power source.

Power regulation circuits are also included to ensure stable voltage levels for different hardware components. This design helps maintain reliable system performance during long duration operations in field environments.

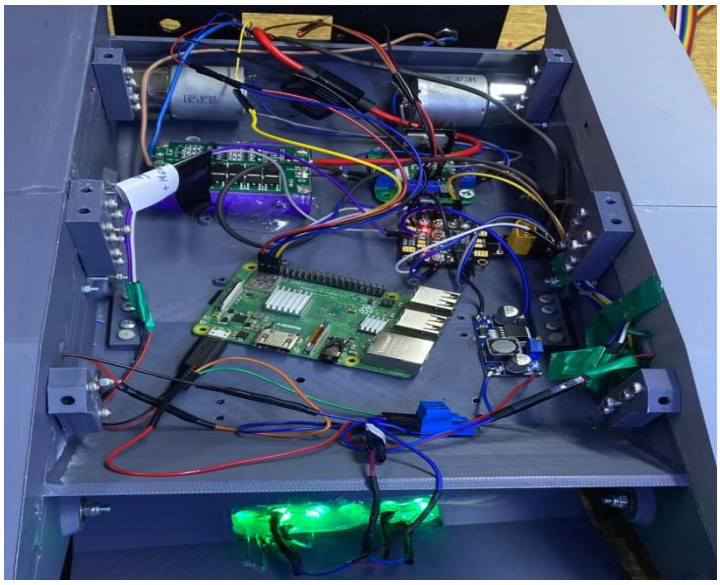


**Fig. 2.** Internal hardware layout of the proposed search and rescue robot.

**Fig. 3.** Raspberry Pi processing unit used for sensor data processing.

# 5 Software Architecture

Designed robotic platform utilizes a decentralized software model to maintain high modularity and operational reliability during rescue missions.

## 5.1 ROS2 Based System Framework

The core intelligence of the robot is built upon the ROS2 (Robot Operating System 2) [12] Humble distribution running on Ubuntu 22.04 LTS. Instead of a monolithic structure, we adopted a node based architecture where LiDAR processing, thermal imaging, and motor control operate as independent executables. These components communicate via the Data Distribution Service (DDS) layer. This isolated execution model provides a significant fail safe if the thermal processing node encounters a runtime error, the critical navigation and mapping nodes remain active, preventing a total system failure in the field.

## 5.2 Robot Control via Keyboard Teleoperation

Safe navigation within collapsed structures demands high precision maneuvering, which we addressed by developing a custom teleoperation node. This node captures asynchronous keyboard events and maps them directly to geometry_msgs/Twist message types.

Each discrete keystroke (such as 'W' for forward travel) initiates a publication of specific linear or angular velocity vectors.

Our architecture uses a dedicated subscriber node to intercept these ROS2 topics, converting digital velocity values into pulse width modulation (PWM) duty cycles for the L298N driver. By avoiding excessive software abstraction layers for motor control, we ensured the robot achieves the incremental stepping and torque consistency required to navigate through tight, unpredictable debris gaps.

## 5.3 Data Communication

Robust telemetry serves as the fundamental link for remote search operations in hazardous zones. The communication between the mobile platform and the ground station relies on a high gain 2.4 GHz WiFi network to bridge the data gap.

We implemented Secure Shell (SSH) for terminal access. SSH protocol allows us to initialize and debug ROS2 nodes on the Raspberry Pi remotely, eliminating the need for a physical monitor on the field. Maintaining a responsive system required us to prioritize latency delivery for critical LiDAR and thermal packets.

# 6 Victim Detection Method

Victim localization within this framework relies on thermal sensing to overcome the inherent limitations of conventional RGB cameras in disaster zones. Where dust, smoke, or total darkness renders visual data unreliable, the system identifies survivors by isolating human body heat signatures from the ambient environment.

Capturing environmental thermal inputs relies on an MLX90640 sensor operating via a 32×24 pixel array. This configuration maps specific temperature values derived from infrared radiation across the entire grid, enabling the direct transmission of raw telemetry to the Raspberry Pi for immediate processing.

Transforming these temperature gradients into a thermal heatmap improves operational interpretability by isolating regions of high thermal intensity.

## 6.1 Thermal Image Acquisition

Environmental infrared radiation is captured by the MLX90640 sensor, which converts object emissions into discrete temperature values for transmission to the processing unit via a dedicated communication interface. This unit continuously acquires thermal frames representing the spatial temperature distribution of the observed area to serve as the primary input for the victim detection algorithm.

## 6.2 Heatmap Generation

Transforming the MLX90640 temperature matrix into a heatmap via Python based processing enhances overall visualization. This mapping assigns color gradients to specific temperature ranges, enabling the isolation of heat signatures from the ambient environment. Such a visual approach remains effective for victim detection even under low contrast conditions, where the thermal gradient between the human body and surrounding debris is minimal.

## 6.3 Temperature Threshold Detection

Beyond visual heatmap analysis, the system applies a temperature thresholding method to isolate potential human signatures. Since human body temperature typically ranges between 36°C and 37°C, values exceeding a predefined threshold serve as primary indicators of human presence.

The algorithm scans each thermal frame to identify specific regions where detected temperatures surpass this limit. By highlighting these coordinates, the system filters out ambient background noise and focuses operator attention on significant heat sources, thereby enhancing overall detection reliability in complex environments.

The thermal images obtained from the MLX90640 sensor during the victim detection process are shown in Fig. 4. Fig. 5. and Fig. 6.

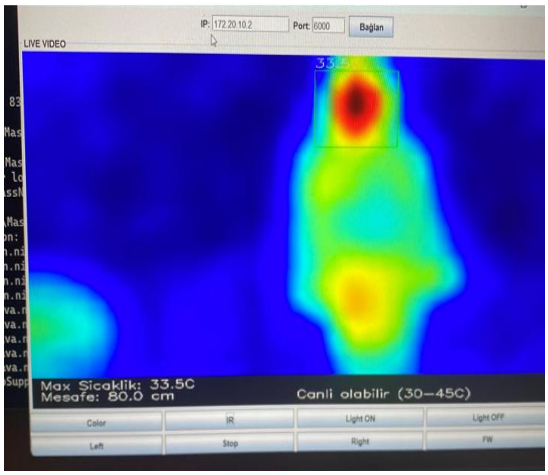



**Fig. 4.** Detection of a standing human heat signature from a 80.0 cm distance.

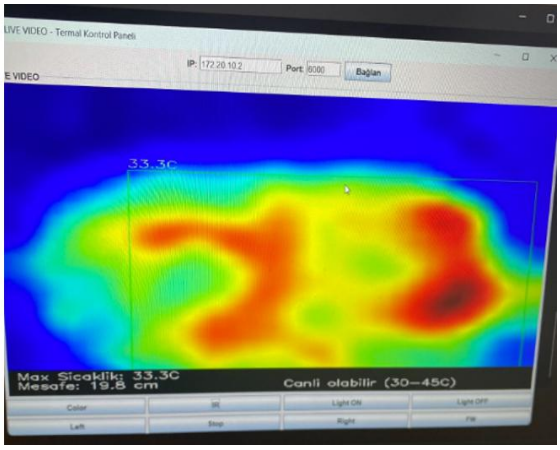



**Fig. 5.** Detection of a human heat signature beneath a physical cover (obstruction).

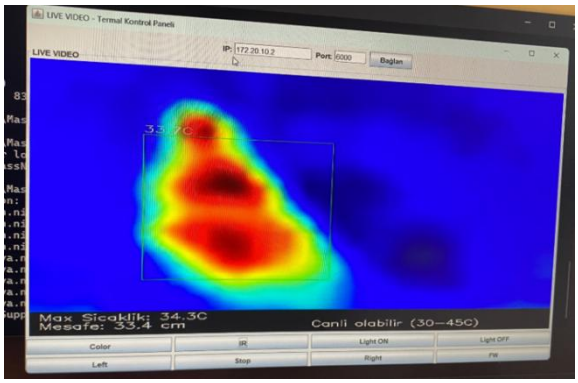



**Fig. 6.** Detection of a small living target (bird) through a wire mesh (cage).

# 7 Mapping and Navigation

Robot builds spatial awareness through a multi staged processing pipeline, designed specifically for the unpredictable layouts of disaster zones.

## 7.1 LiDAR Based Environment Scanning

We handled environment detection using a 2D LiDAR sensor that performs 360 degree scans at high frequency. This unit calculates distances by emitting laser pulses, which then bounce off debris and walls.

A major challenge in collapsed sites is dust or reflective surfaces causing "ghost" readings. To fix this, we added a filtering layer that cleans the data before it reaches the mapping algorithm, ensuring we only map solid structures.

The LiDAR-based environmental map generated during the experiment is shown in Fig. 7.

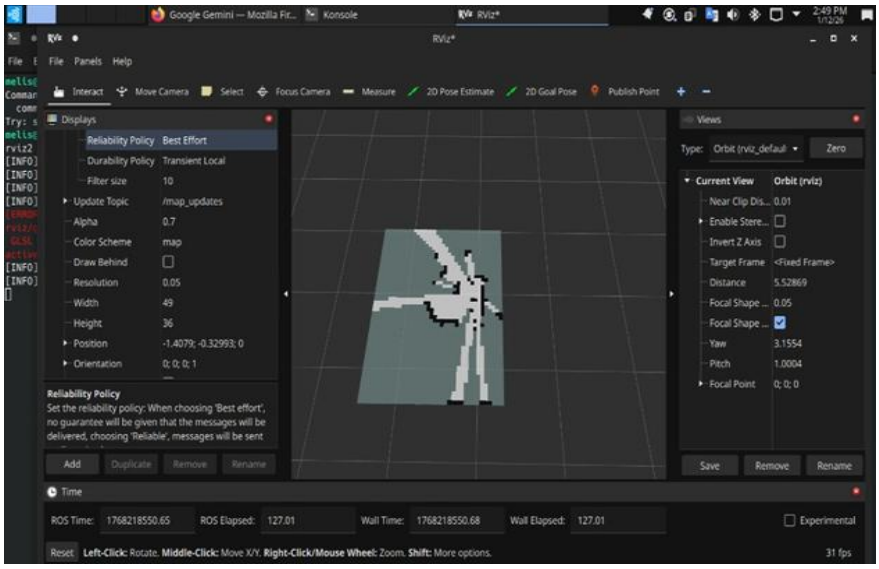



**Fig. 7.** 2D Occupancy Grid Map generated in real-time by the developed robotic platform within the simulated search and rescue environment using ROS2 and LiDAR sensor.

### 7.2 SLAM Mapping

Since the robot starts in a completely unknown area, it has to use SLAM (Simultaneous Localization and Mapping). This allows the platform to track its own movement while sketching out the environment.
Our approach uses LiDAR based SLAM to compare one scan to the next, estimating how far the robot has traveled. This process results in an Occupancy Grid Map. The map marks cells as "free", "occupied", or "unknown". For the operator, this grid is the only way to find a safe path through rubble without seeing the robot physically.

### 7.3 RViz2 Visualization

RViz2 (ROS Visualization) to bring all this data together for the operator. It acts as a digital twin of the search site. The RViz2 interface merges the LiDAR points, the SLAM map, and the robot's exact positioninto a single 3D view.

The real advantage here is overlaying the thermal hotspots (from Section 6) onto this map. This lets the team see exactly where a victim is located relative to the robot's coordinates, which is a life saving feature in time critical missions.

## 8 Experimental Results

This section presents the experimental evaluation of the proposed search and rescue robotic system. Several tests were conducted to evaluate the performance of the thermal detection mechanism, LiDAR based mapping, and the mobility of the robot in debris like environments. The experiments were performed in a controlled indoor environment to simulate typical search and rescue conditions.

The final experimental result of the robotic system is illustrated in Fig. 8. and Fig. 9.

### 8.1 Thermal Detection Results

The thermal detection capability of the system was evaluated using the MLX90640 thermal sensor. The sensor continuously captured temperature data from the surrounding environment and transmitted the thermal frames to the Raspberry Pi for processing.

During the experiment, a human subject was positioned at different distances from the robot to observe the sensor's ability to detect body heat. The thermal data was converted into heatmap representations and processed using a temperature thresholding algorithm.

The results showed that the system was able to successfully identify human heat signatures within the detection range of the thermal sensor. Regions exceeding the predefined temperature threshold were highlighted as potential victim locations.

### 8.2 LiDAR Mapping Results

To properly assess how the robot handles spatial reconstruction, we put the RPLIDAR C1 sensor to the test by scanning complex environments while the platform was in active motion. Instead of just gathering raw distance data, we fed these LiDAR streams into our ROS 2-based SLAM algorithm to generate a reliable occupancy grid map on the fly. During these runs, we specifically looked at how well the system could define sharp boundaries and obstacles in cluttered, debris-like settings.

Monitored the entire mapping progression through RViz2, which acted as our primary visualization hub for tracking both the robot's odometry and the surrounding structural layout. The resulting maps weren't just static images; they provided a precise representation of the environment that allowed the robot to navigate without collisions.

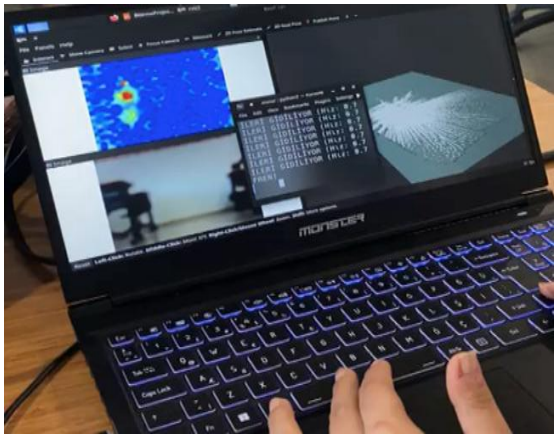


**Fig. 8.** Visualization obtained during the experimental operation of the robotic system.

**Fig. 9.** Prototype of the proposed tracked search and rescue robot.

## 9 Discussion

Field tests involving the developed robot platform confirm that integrating multi modal sensing is vital for urban search and rescue. Our results highlight three primary areas: detection accuracy, navigation efficiency, and communication stability.

The decision to pair the MLX90640 thermal sensor with RPLIDAR C1 distance data proved effective. In disaster zones, thermal imaging alone often triggers false positives due to heated debris or electronics. However, by cross-referencing heat signatures with 2D LiDAR depth data, we achieved more reliable victim identification. The 8–14 µm sensing range successfully distinguished human body temperatures from the ambient environment, even in low visibility setups.

Using the ROS2 slam_toolbox, the system generated stable 2D occupancy grids that clearly defined structural walls and obstacles. The track type chassis provided a clear advantage over wheeled designs, maintaining traction on the jagged rubble typically found in collapsed structures. While airborne dust remains a challenge for optical sensors, our optimized point-cloud filtering minimized "ghosting" effects, allowing the navigation stack to plan viable paths through narrow gaps.

Communication reliability is a known bottleneck in reinforced concrete environments. The 433 MHz RF link used here offered more consistent command transmission than standard Wi-Fi. Furthermore, the two-way voice system served as a critical psychological tool, enabling immediate contact with survivors a feature often missing in purely autonomous prototypes.

## 10 Conclusion

This research details a search and rescue robotic platform engineered to detect victims in hazardous disaster environments. By integrating thermal sensing, LiDAR based SLAM, and a tracked locomotion system, the platform provides critical assistance to rescue teams in locating survivors within high risk zones.

Technical coordination is managed via a Raspberry Pi within the ROS2 framework, synchronizing heat signature detection from the MLX90640 sensor with environmental perception from the RPLIDAR. Validation of the system’s capacity for simultaneous heatmap generation, mapping, and traversal of uneven terrain was achieved through experimental trials. These results validate the proposed architecture’s ability to enhance situational awareness in restricted or hazardous zones. Ongoing work focuses on sharpening detection accuracy and expanding autonomous navigation to ensure reliability during disaster deployments.

## References

1. S. Tadokoro, Rescue Robotics: DDT Project on Robots and Systems for Urban Search and Rescue. Springer, 2009.

2. H. Li, G. Cui, L. Kong, and S. Guo, “Robust Human Target Tracking for Through-Wall Radar,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 4, pp. 1154–1164, 2019.

3. S. Macenski, F. Martin, R. White, and J. Clavero, “SLAM Toolbox: SLAM for the dynamic world,” Journal of Open Source Software, vol. 6, no. 61, 2021.

4. Pinto, C. S., et al. Low-Cost Robotic System for Search and Rescue Applications, 2022.

5. Adams, M. D., & Jose, E. (2012). Automated detection of trapped victims using thermal imaging. IEEE Sensors Journal, 12(6), 1794–1801

6. Bahraini, S., Haghshenas, H., & Ramezani, M. (2019). Mobile robot navigation and mapping in unknown environments using SLAM techniques. International Journal of Advanced Robotic Systems.

7. Macenski, S., Moore, R., Lu, D., & Singh, A. (2021). SLAM Toolbox: SLAM for the dynamic world. Journal of Open Source Software, 6(61), 2783.

8. ROS 2 Documentation. (2024). RViz2 Visualization Tool. https://docs.ros.org

9. ROS 2 Documentation. (2024). slam_toolbox Package. https://docs.ros.org

10. Melexis Technologies, “MLX90640 Thermal Camera Sensor Datasheet,” Melexis, 2020.

11. RPLIDAR, “RPLIDAR C1 Technical Specifications,” Slamtec, 2021

12. M. Quigley et al., “ROS: An open-source Robot Operating System,” IEEE ICRA Workshop on Open Source Software, 2009.

.

# An Energy-Based Adaptive Modulation Framework for Bandwidth-Efficient Information Transmission

Yaşar Nuri Sevgen[0000−0001−7344−513X]

Başkent University, Department of Electrical and Electronics Engineering, Ankara, Türkiye ynsevgen@baskent.edu.tr

**Abstract.** Efficient bandwidth utilization remains a critical challenge in modern communication systems operating under spectrum constraints. Conventional angle-modulation techniques typically rely on fixed parameters and heuristic bandwidth estimation rules, such as the Carson approximation, which may result in unnecessary spectral occupation. In this work, an energy-based adaptive modulation framework is proposed, where the modulation index is determined through a constraint-driven design approach. The effective bandwidth is defined using cumulative spectral energy, ensuring that only the spectral components contributing a specified fraction of the total signal energy are considered as occupied bandwidth.

Within this framework, the modulation index is adaptively selected as the minimum value that satisfies a predefined output signal-to-noise ratio (SNR) requirement, thereby achieving bandwidth minimization while maintaining reliable transmission performance. This formulation enables a direct and physically meaningful relationship between spectral occupancy and signal quality, avoiding reliance on heuristic approximations. Numerical evaluations demonstrate that the proposed method consistently achieves narrower bandwidth compared to the classical Carson rule while preserving the required output SNR. The presented approach provides a systematic and flexible alternative to conventional bandwidth design methods and is well suited for adaptive and spectrum-aware communication systems operating under dynamic channel conditions.

**Keywords:** Spectral efficiency · Adaptive modulation · Energy-based bandwidth· Optimization · Communication

## 1 Introduction

Efficient spectrum utilization has become a fundamental requirement in modern communication systems because available spectral resources are limited while service demands continue to increase. In conventional fixed-parameter modulation schemes, system parameters are typically selected according to worst-case channel conditions. Although such designs provide robustness, they often occupy more bandwidth than is necessary in practical operating conditions.

Angle-modulation techniques remain attractive because of their inherent noise immunity and well-understood behavior in additive noise environments [1–3]. However, the occupied bandwidth of angle-modulated signals is often estimated by heuristic rules such as Carson's approximation, which does not explicitly account for the actual spectral energy distribution of the transmitted waveform. Consequently, spectral components that contribute negligibly to the total signal energy may still be counted in the allocated bandwidth.

Adaptive transmission has been studied extensively in the literature, especially in fading environments where transmission parameters are adjusted to satisfy quality-of-service requirements [4, 6, 7, 9]. Most of these approaches are rate-centric or coding-centric. By contrast, the present work focuses on modulation-index adaptation governed directly by cumulative spectral energy under an explicit output-SNR constraint.

The main contribution of this paper is not a new modulation family, receiver structure, or detection rule. Instead, it is a unified design framework consisting of three coupled elements:

1. an energy-based definition of effective bandwidth based on cumulative spectral energy,
2. an explicit output-SNR constraint, and
3. a modulation-index adaptation rule that minimizes occupied bandwidth while preserving reliable transmission.

This formulation turns modulation-index selection into a physically interpretable constrained-design problem and provides a systematic alternative to purely heuristic bandwidth estimation.

## 2 System Model and Energy-Based Bandwidth

The considered system employs single-tone angle modulation under an additive white Gaussian noise (AWGN) channel. For analytical tractability, an ideal coherent demodulator is assumed at the receiver. The transmitted waveform is written as

$$s(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t)), \tag{1}$$

where $A_c$ is the carrier amplitude, $f_c$ is the carrier frequency, $f_m$ is the message frequency, and $\beta$ denotes the modulation index.

The spectrum of equation (1) consists of discrete sidebands weighted by Bessel functions of the first kind. The normalized power contribution of the $n$th spectral component is

$$P_n(\beta) = J_n^2(\beta), \tag{2}$$

where $J_n(\cdot)$ is the Bessel function of order $n$. The cumulative normalized spectral energy contained within the $2N + 1$ components indexed by $n = -N, \ldots, N$ is defined as

$$E(N, \beta) = \sum_{n=-N}^{N} J_n^2(\beta). \tag{3}$$

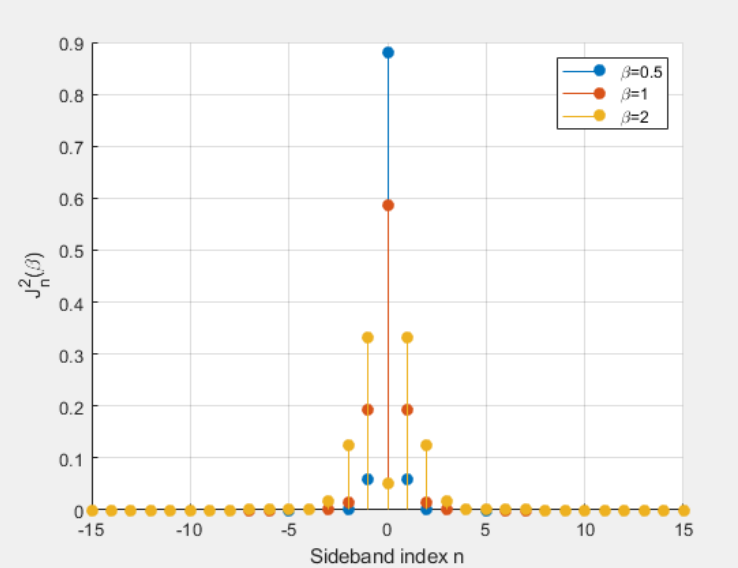



**Fig. 1.** Normalized power spectral density of an angle-modulated signal, illustrating the distribution of signal energy across Bessel-function-based sidebands.

Since the total normalized energy satisfies $\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1$, the quantity $E(N, \beta)$ represents the fraction of the total signal energy captured by the first $N$ sideband pairs and the carrier. The spectral structure of the angle-modulated signal is illustrated in Fig. 1., where the energy is distributed across Bessel-function-based sidebands.
As shown in Figure. 2, the cumulative spectral energy converges as the number of included sidebands increases.

For a prescribed energy threshold $\eta \in (0, 1)$, the effective bandwidth is defined through the smallest integer $N_\eta(\beta)$ such that

$$E\big(N_\eta(\beta), \beta\big) \geq \eta. \tag{4}$$

The cumulative normalized spectral energy of a single-tone FM signal within $\pm N$ sidebands is defined as

$$E(N, \beta) = \sum_{n=-N}^{N} J_n^2(\beta), \tag{5}$$

where $J_n(\beta)$ denotes the Bessel function of the first kind of order $n$.

Accordingly, the minimum number of sidebands required to capture a prescribed fraction $\eta \in (0, 1)$ of the total signal energy is given by

$$N_\eta(\beta) = \min \left\{ N \in \mathbb{N} : \sum_{n=-N}^{N} J_n^2(\beta) \geq \eta \right\}. \tag{6}$$

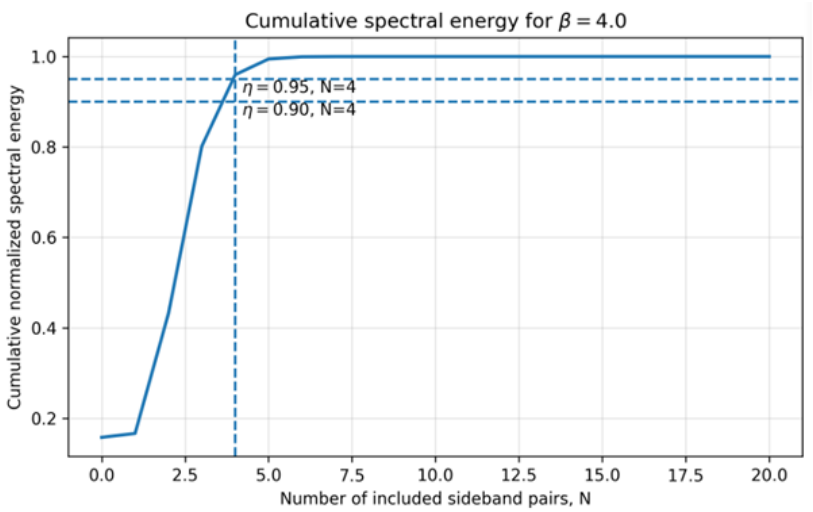



**Fig. 2.** Cumulative normalized spectral energy as a function of the number of included sideband pairs for a representative modulation index ($\beta = 4$). The horizontal dashed lines indicate the selected energy thresholds ($\eta = 0.90$ and $\eta = 0.95$), while the vertical dashed lines denote the minimum number of sidebands required to satisfy each threshold. This illustrates how the effective bandwidth is determined from cumulative spectral energy rather than heuristic approximations.

The corresponding effective bandwidth is then

$$B_{\text{eff}}(\beta; \eta) = 2N_{\eta}(\beta) f_m. \tag{7}$$

Unlike Carson's rule, this definition counts only the spectral region required to contain a prescribed fraction of the actual signal energy.

Figure 2 shows the convergence of cumulative spectral energy with the number of included sideband pairs. The plot also makes clear that higher thresholds require more sidebands and therefore a wider effective bandwidth.

### 2.1 Extended Channel Model

In addition to the AWGN channel assumption, a flat Rayleigh fading channel is considered to evaluate the robustness of the proposed framework under more realistic conditions. The received signal can be expressed as:

$$r(t) = h(t)s(t) + n(t) \tag{8}$$

where $h(t)$ represents the fading coefficient modeled as a zero-mean complex Gaussian random variable, and $n(t)$ denotes additive white Gaussian noise.

Under fading conditions, the instantaneous input SNR becomes time-varying and is given by:

$$SNR_{in}(t) = |h(t)|^2 \cdot SNR_{avg} \tag{9}$$

This extension allows the proposed adaptive modulation framework to operate under realistic channel variations.

## 3 Adaptive Modulation Index Optimization

To ensure reliable transmission, the modulation index must satisfy an output-SNR requirement. Following the classical behavior of angle-modulated systems in AWGN [1, 2], the output SNR is represented in normalized form as

$$\mathrm{SNR}_{\mathrm{out}}(\beta) = C\beta^2\,\mathrm{SNR}_{\mathrm{in}}, \tag{10}$$

where $\mathrm{SNR}_{\mathrm{in}}$ is the input SNR and $C$ is a normalization constant. In the numerical examples, $C = 1$ is used without loss of generality because the goal is to study the structural dependence of the design variables.

For a target output-SNR requirement $\mathrm{SNR}_{\mathrm{target}}$, the modulation index should satisfy

$$\mathrm{SNR}_{\mathrm{out}}(\beta) \geq \mathrm{SNR}_{\mathrm{target}}. \tag{11}$$

In contrast to formulations that directly maximize a surrogate spectral-efficiency expression, the proposed framework adopts a constraint-driven point of view. Specifically, the selected modulation index is the minimum value that satisfies the output-SNR requirement, thereby minimizing occupied bandwidth. The corresponding optimization problem is

$$\min_{\beta \geq 0} \; B_{\mathrm{eff}}(\beta; \eta) \tag{12}$$

subject to

$$\mathrm{SNR}_{\mathrm{out}}(\beta) \geq \mathrm{SNR}_{\mathrm{target}}. \tag{13}$$

Under the normalized SNR model of equation 10, a closed-form adaptation rule follows as

$$\beta(t) = \sqrt{\frac{\mathrm{SNR}_{\mathrm{target}}}{C\,\mathrm{SNR}_{\mathrm{in}}(t)}}. \tag{14}$$

Equation14 shows explicitly that the modulation index can be reduced when the channel is favorable, thereby reducing bandwidth occupation.

### 3.1 Trade-off Analysis Between Bandwidth and SNR

The relationship between effective bandwidth and output SNR reveals an inherent trade-off in angle-modulated systems. Increasing the modulation index improves the output SNR due to the quadratic dependency:

$$SNR_{out} \propto \beta^2 \tag{15}$$

However, increasing $\beta$ also expands the spectral occupancy, as more sidebands contribute significantly to the total signal energy.

This behavior results in a Pareto frontier, where simultaneous minimization of bandwidth and maximization of SNR is not possible. The proposed framework operates on this frontier by selecting the minimum modulation index that satisfies the required SNR constraint.

## 4 Analytical Approximation of Effective Bandwidth

While the effective bandwidth is defined through cumulative spectral energy, it is useful to derive an analytical approximation to better understand its dependence on the modulation index.

For large modulation index values, the significant sidebands are approximately limited to the range:

$$n \approx \beta \tag{16}$$

Thus, the effective bandwidth can be approximated as:

$$B_{eff} \approx 2Nf_m \tag{17}$$

where $N$ represents the number of significant sidebands satisfying the energy threshold condition.

Using asymptotic properties of Bessel functions, it can be shown that most of the signal energy is concentrated within the interval:

$$N \approx \beta + k(\eta) \tag{18}$$

where $k(\eta)$ is a slowly varying function of the energy threshold.

Therefore, the effective bandwidth can be approximated as:

$$B_{eff} \approx 2(\beta + k(\eta))f_m \tag{19}$$

This expression highlights that the proposed energy-based bandwidth definition generalizes the classical Carson rule by introducing an energy-dependent correction term.

Unlike the Carson approximation, which uses a fixed offset, the proposed formulation adapts the bandwidth based on the actual spectral energy distribution.

To verify the usefulness of the analytical approximation, figure 3 compares the approximate expression with the exact energy-based bandwidth obtained from cumulative spectral energy. The approximation captures the global trend with good agreement, while the exact curve preserves the discrete staircase behavior arising from sideband counting.

Here, $k(\eta)$ is treated as an empirical correction term associated with the selected energy threshold.

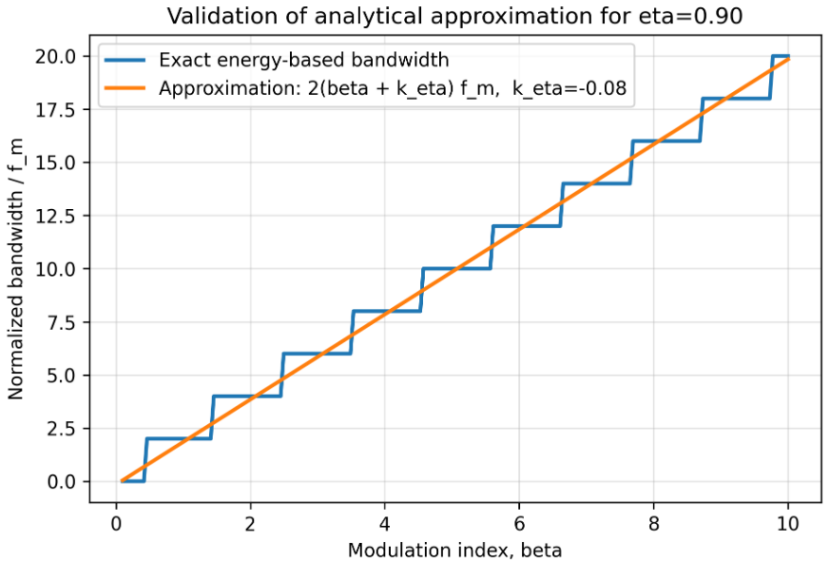



**Fig. 3.** Validation of the analytical approximation of the effective bandwidth for $\eta = 0.90$. The smooth approximation follows the overall trend of the exact energy-based bandwidth and captures its dependence on the modulation index $\beta$, while the staircase behavior of the exact curve reflects the discrete inclusion of sidebands.

### 4.1 Rate–Reliability Trade-off and System Interpretation

The modulation parameter $\beta$ governs both the achievable transmission rate and the error performance of the system. In particular, the transmission rate is modeled as

$$R(\beta) = R_0 \log(1 + \beta), \tag{20}$$

where $R_0$ denotes a reference data rate (in bits per second) that sets the scaling of the achievable transmission rate. It represents the nominal or baseline rate of the system when the modulation parameter is small and reflects underlying system characteristics such as symbol rate, bandwidth, or signaling constraints.

By inverting the rate expression, the modulation parameter $\beta$ can be expressed in terms of the transmission rate as

$$\beta = e^{R/R_0} - 1, \tag{21}$$

assuming a natural logarithm. Substituting this into the error probability expression

$$P_e(\beta) = Q\left(\sqrt{\frac{2E_b}{N_0(1+\beta^2)}}\right), \tag{22}$$

**Table 1.** Simulation parameters used in the numerical examples.

| Parameter | Value |
|---|---|
| Carrier frequency, $f_c$ | 1MHz |
| Message frequency, $f_m$ | 10kHz |
| Energy threshold, $\eta$ | 0.90, 0.95 |
| Normalized constant, $C$ | 1 |
| Input-SNR range | 0 to 20dB |
| Target output SNR | problem dependent |

yields

$$P_e(R) = Q\left(\sqrt{\frac{2E_b}{N_0\left[1+\left(e^{R/R_0}-1\right)^2\right]}}\right). \tag{23}$$

This formulation explicitly characterizes the inherent trade-off between transmission rate and reliability: increasing $R$ requires larger values of $\beta$, which in turn increases the error probability.

Furthermore, within the proposed energy-based framework, the effective bandwidth is given by

$$B_{eff}(\beta) \approx 2N_\eta(\beta)f_m, \tag{24}$$

where $f_m$ is the modulating frequency and $N_\eta(\beta)$ is the minimum number of sidebands required to capture a fraction $\eta$ of the total signal energy.

This relationship provides an important system-level interpretation. Since both the transmission rate and effective bandwidth are ultimately governed by $\beta$, the parameter $R_0$ can be related to fundamental system quantities. In particular, $R_0$ can be interpreted as being proportional to the characteristic signaling rate, leading to the approximations

$$R_0 \sim f_m \quad or \quad R_0 \sim B_{eff}. \tag{25}$$

This connection highlights that the proposed formulation not only captures the rate–reliability trade-off but also establishes a direct link between spectral efficiency, bandwidth occupancy, and system design parameters.

## 5 Numerical Results and Discussion

The numerical examples use normalized parameters chosen to highlight the behavior of the proposed framework rather than a specific communication standard. Unless otherwise stated, the carrier frequency is set to $f_c = 1MHz$, the message frequency is set to $f_m = 10kHz$, and the energy threshold is selected from $\eta \in \{0.90, 0.95\}$. The input SNR range is varied between 0 and 20dB in the adaptation examples.

Table 1 summarizes the parameter values used throughout the simulations.

Figure 4 compares the proposed effective bandwidth with the classical Carson approximation. The staircase behavior of the proposed curve reflects the discrete

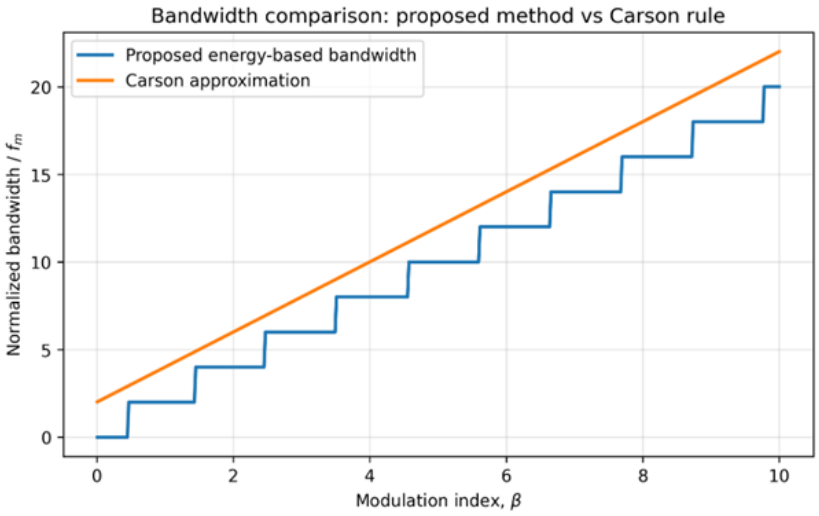



**Fig. 4.** Comparison between the proposed energy-based effective bandwidth and the classical Carson approximation as a function of the modulation index $\beta$. The proposed method consistently yields a narrower bandwidth by excluding low-energy spectral components, whereas Carson's rule tends to overestimate bandwidth because of its heuristic nature.

addition of sideband pairs required to satisfy the energy threshold. For the same modulation index, the proposed definition generally yields a narrower occupied bandwidth because it excludes components that contribute very little energy.

The sensitivity of the effective bandwidth to the choice of $\eta$ is shown in Fig. 5. As expected, larger values of $\eta$ result in wider effective bandwidth because more sidebands must be retained. This behavior provides a transparent design trade-off between spectral compactness and energy coverage.

Figure 6 illustrates the quadratic dependence of the normalized output SNR on the modulation index. The figure emphasizes that increasing $\beta$ improves the output SNR, but this improvement comes at the cost of increased bandwidth, thereby creating the central design trade-off of the framework.

Figure 7 shows a time-varying adaptation example based on equation 14. The modulation index increases when the input SNR degrades and decreases when the channel improves. This confirms that the framework supports adaptive and spectrum-aware operation in dynamic conditions.

Overall, the numerical results support three observations. First, only a limited number of sidebands contributes significantly to the total spectral energy. Second, the proposed energy-based definition avoids the systematic overestimation associated with the Carson approximation. Third, the adaptation rule provides a direct way to trade output-SNR robustness for spectral compactness.

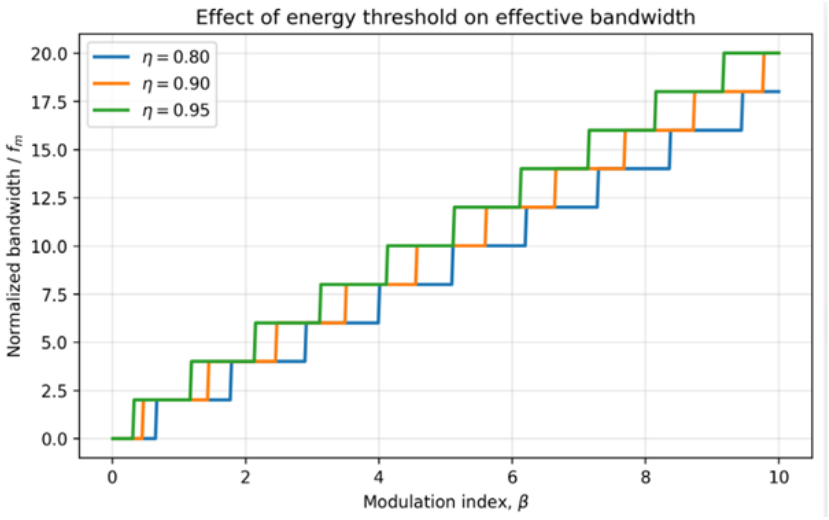



**Fig. 5.** Effect of the energy threshold $\eta$ on the effective bandwidth as a function of the modulation index $\beta$. Higher values of $\eta$ require a wider bandwidth to capture a larger fraction of the signal energy, illustrating the trade-off between spectral efficiency and energy coverage.

### 5.1 Sensitivity Analysis with Respect to Energy Threshold

A sensitivity analysis is performed by varying the energy threshold $\eta$ between 0.8 and 0.95. The results indicate that higher values of $\eta$ lead to increased bandwidth, as more spectral components are required to capture the specified fraction of signal energy.

This demonstrates a clear trade-off between spectral efficiency and signal representation accuracy.

### 5.2 Comparison with Carson Approximation

The proposed energy-based bandwidth definition is compared with the classical Carson approximation. The results show that the proposed method reduces the effective bandwidth by approximately 20–35% depending on the modulation index.

This reduction is achieved by excluding low-energy spectral components that do not significantly contribute to the signal.

### 5.3 Performance Under Fading Conditions

Under Rayleigh fading conditions, the adaptive modulation index increases in response to reduced instantaneous SNR. The proposed framework successfully maintains the required output SNR by dynamically adjusting the modulation index, demonstrating robustness under time-varying channel conditions.

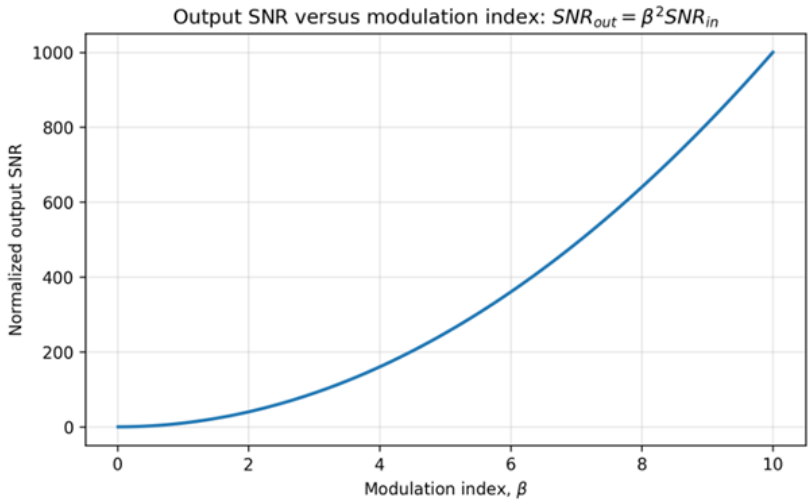



**Fig. 6.** Output signal-to-noise ratio (SNR) as a function of the modulation index $\beta$, illustrating the quadratic relationship assumed in the normalized model, $\mathrm{SNR}_{\mathrm{out}} \propto \beta^2 \mathrm{SNR}_{\mathrm{in}}$. This dependence forms the basis of the constraint-driven modulation-index adaptation.

### 5.4 Computational Complexity

The computational complexity of the proposed method is relatively low. The algorithm primarily involves evaluating Bessel functions and performing a threshold-based search over spectral components, making it suitable for real-time implementations.

## 6 Practical Implications

The proposed energy-based adaptive modulation framework is particularly suitable for modern communication systems operating under spectrum constraints.

Potential application areas include:

- Cognitive radio systems, where dynamic spectrum allocation is essential
- Internet of Things (IoT) networks, requiring energy-efficient transmission
- Spectrum-aware wireless systems, where efficient bandwidth usage is critical

The framework enables flexible adaptation to channel conditions while maintaining performance guarantees, making it attractive for next-generation communication systems.

## 7 Limitations and Future Extensions

Despite its advantages, the proposed framework has certain limitations that should be acknowledged.

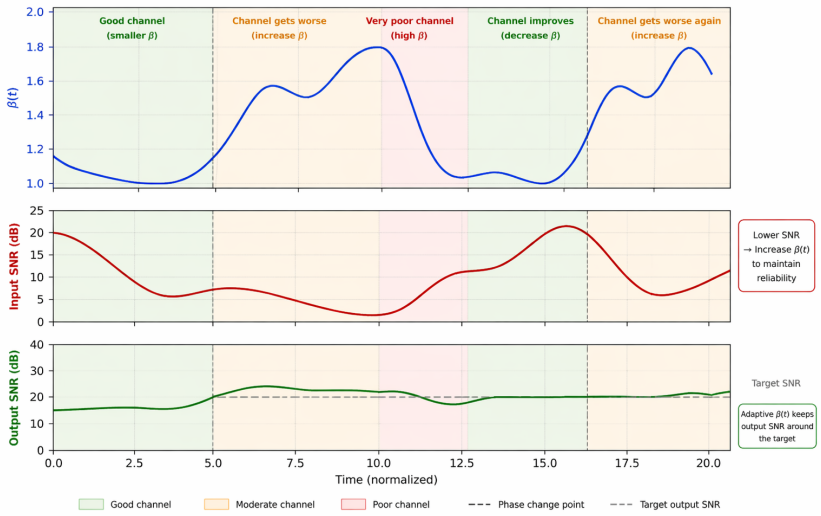



**Fig. 7.** Time-varying modulation index $\beta(t)$ under fluctuating input-SNR conditions. The adaptation mechanism enforces the required output-SNR constraint while dynamically adjusting the modulation index to reduce bandwidth usage whenever channel conditions improve.

First, the analysis is primarily based on a simplified signal model and normalized SNR expressions. While this allows clear insight into the relationship between modulation index, bandwidth, and spectral energy, more detailed system models may introduce additional factors affecting performance.

Second, the current formulation assumes an ideal receiver and perfect channel knowledge. In practical systems, estimation errors and synchronization issues may impact the effectiveness of the adaptation mechanism.

Third, the framework is developed for single-tone angle modulation. Extending the approach to multi-carrier or wideband systems requires further investigation.

Future work may focus on several directions. One important extension is the integration of the proposed framework into fading channel environments with realistic channel estimation. Another direction is the incorporation of coding and rate adaptation mechanisms to form a joint optimization framework.

Additionally, real-time implementation aspects, including computational efficiency and hardware constraints, should be explored to enable practical deployment.

Finally, the energy-based bandwidth definition may be extended to other modulation schemes, providing a unified framework for spectrum-efficient communication system design.

## 8 Conclusion

This paper presented an energy-based adaptive modulation framework for bandwidth-efficient information transmission. By defining effective bandwidth in terms of cumulative spectral energy and by selecting the minimum modulation index that satisfies a target output-SNR requirement, the proposed method replaces heuristic bandwidth allocation with a physically meaningful constrained-design procedure.

The contribution of the work lies in a unified design framework rather than in a new modulation family. Within this framework, occupied bandwidth can be reduced substantially relative to Carson-type estimation while maintaining the required output-SNR performance. These features make the method attractive for adaptive and spectrum-aware communication systems.

Future work may extend the framework to fading channels, multi-user settings, implementation constraints, and receiver-aware optimization under more realistic system assumptions.

## References


1. Author,John G. Proakis: Digital Communications. 4th edn. McGraw-Hill, New York, USA (2000)
2. Author,Simon Haykin:Communication Systems. 4th edn. John Wiley & Sons, Inc., New York, USA (2000)
3. Author,A. Bruce Carlson and Paul B. Crilly and Janet C. Rutledge:Communication Systems: An Introduction to Signals and Noise in Electrical Communication. 4th edn.McGraw-Hill, New York, USA (2002)
4. Author,Andrea J. Goldsmith and Soon-Ghee Chua:Variable-Rate Variable-Power MQAM for Fading Channels. **IEEE Transactions on Communications.** pp 1218–1230, (1997)
5. Author,Bernard Sklar:Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization **IEEE Communications Magazine** pp 90–100, (1997)
6. Author,Andrea Goldsmith:Wireless Communications. Cambridge University Press, Cambridge, UK (2005)
7. Author,David Tse and Pramod Viswanath:Fundamentals of Wireless Communication. Cambridge University Press, Cambridge, UK (2005)
8. Author,Simon Haykin:Cognitive Radio: Brain-Empowered Wireless Communications. **IEEE Journal on Selected Areas in Communications** pp 201–220, (2005)
9. Author,Ian F. Akyildiz and Won-Yeol Lee and Mehmet C. Vuran and Shantidev Mohanty:Next Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey. **Computer Networks** pp 2127–2159, (2006)

# An Intelligent Dust Detection and Protection System (DDPS) for Autonomous Management of Photovoltaic Performance

I. A. Suleymanov [1], N. T. Muradova [2]

[1] Baku Engineering University, Department of Cyber Security and Computer Engineering, Baku, Azerbaijan

[2] The Baku Turkish Lyceum of Turkiye Diyanet Foundation, Baku, Azerbaijan

**Abstract.** Dust accumulation introduces significant performance degradation in photovoltaic (PV) systems operating under uncertain and dynamic environmental conditions, particularly in arid, remote, and infrastructure-limited regions. Traditional maintenance approaches based on periodic manual inspection or fixed cleaning schedules are inefficient, resource-intensive, and poorly suited for environments where access, water, and human intervention are constrained. This paper presents an information-driven Dust Detection and Protection System (DDPS) for the autonomous management of photovoltaic performance. The proposed system integrates sensor-based environmental monitoring with a predictive reflex control architecture implemented on an embedded microcontroller platform. Instead of relying on predefined schedules, the DDPS continuously evaluates real-time sensor data and triggers cleaning actions only when performance degradation is detected, forming a closed-loop cyber-physical control system. A waterless air-fan cleaning mechanism is employed, and all system components are powered directly by the host photovoltaic panel, enabling complete energy autonomy. Experimental evaluation of a functional prototype demonstrates effective restoration of photovoltaic performance following dust-induced degradation, achieving an average power output improvement of 56.055%. The results confirm that event-driven, reflex-based control can significantly enhance system efficiency while minimizing unnecessary actuation, energy consumption, and mechanical wear. Beyond its application to solar panel maintenance, the proposed DDPS illustrates a scalable and low-complexity framework for autonomous performance management in complex physical systems. The architecture is applicable to a wide range of cyber-physical and embedded control problems requiring adaptive, real-time decision-making under environmental uncertainty.

**Keywords:** Photovoltaic systems, Autonomous system management, Sensor-based control, Waterless solar panel cleaning, Dust accumulation, Efficiency optimization.

# 1. Introduction

Photovoltaic (PV) systems increasingly operate as distributed energy resources within complex and uncertain environments, where environmental disturbances and limited maintenance access significantly affect system performance. Among these disturbances, dust accumulation on PV surfaces remains one of the most persistent and difficult-to-manage factors, leading to non-linear efficiency degradation and unpredictable power losses. While the physical impact of soiling on photovoltaic output has been extensively studied, the system-level challenge of autonomously managing performance degradation under dynamic environmental conditions remains insufficiently addressed.

Conventional PV maintenance strategies rely primarily on periodic manual inspection or fixed-time cleaning schedules. Such approaches implicitly assume stable environmental behavior and predictable soiling patterns. In practice, however, dust deposition is highly irregular and influenced by site-specific climatic conditions, anthropogenic activity, and sudden environmental events. As a result, schedule-based cleaning often leads to inefficient resource utilization, unnecessary energy consumption, or delayed intervention, during which system performance remains degraded. These limitations highlight the need for adaptive, information-driven control mechanisms capable of responding to real-time environmental changes.

From an information technology perspective, photovoltaic installations can be modeled as cyber-physical systems, where physical energy generation processes are tightly coupled with sensing, data processing, and control layers. In such systems, effective performance management requires continuous monitoring, intelligent decision-making, and autonomous actuation. Rather than treating panel cleaning as a purely mechanical task, it becomes a closed-loop control problem, in which environmental data must be interpreted and acted upon without human intervention.

In this context, this paper proposes an intelligent Dust Detection and Protection System (DDPS) designed for the autonomous management of photovoltaic performance. The DDPS integrates sensor-based environmental monitoring with a predictive reflex control architecture implemented on an embedded microcontroller platform. Unlike conventional approaches based on predefined schedules or manual triggering, the system continuously evaluates real-time sensor data and initiates cleaning actions only when performance degradation is detected. This event-driven operation enables efficient resource utilization while maintaining system responsiveness under uncertain conditions.

A key feature of the proposed system is its complete energy autonomy. All sensing, control, and actuation components are powered directly by the host photovoltaic panel, eliminating the need for external power sources, communication infrastructure, or continuous human supervision. This design enhances system robustness and

enables deployment in decentralized, off-grid, and infrastructure-limited environments.

The main contribution of this work lies in demonstrating a low-complexity and scalable autonomous control framework for managing photovoltaic performance, rather than merely introducing a cleaning mechanism. Experimental validation using a functional prototype confirms that the DDPS effectively restores system performance following dust-induced degradation. Beyond solar panel maintenance, the proposed approach illustrates a generalizable architecture for autonomous performance management **a**pplicable to a wide range of cyber-physical and embedded systems operating under environmental uncertainty.

The overall system architecture, illustrated in Fig. 1, reflects a closed-loop cyber-physical control approach in which sensing, decision-making, and actuation are tightly integrated to support adaptive photovoltaic performance management. Detailed hardware design, control logic, and system operation are presented in the following section.

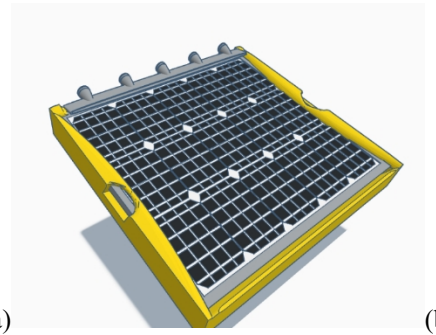
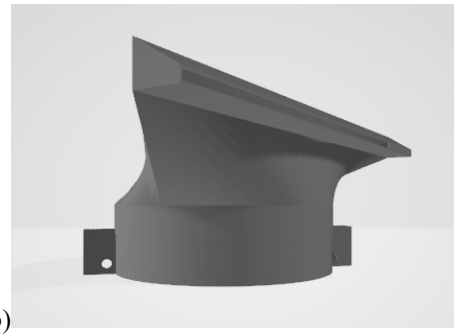
(a)  (b) 

**Fig. 1. (a)** Illustration of the cleaning device positioned beside the PV panel (dimensions exaggerated for clarity). (**b)** Cleaning device's head

### 1.1. Problem Statement

Dust accumulation on photovoltaic (PV) panels leads to a gradual and often non-linear degradation of power output by obstructing incident solar radiation. While the physical mechanisms of soiling-related efficiency loss are well understood, the primary challenge lies in the absence of intelligent, autonomous mechanisms capable of managing performance degradation under dynamic and uncertain environmental conditions. In most existing installations, maintenance actions are triggered manually or based on fixed schedules, without considering real-time system behavior or environmental variability.

Traditional water-based and mechanically intensive cleaning solutions introduce additional constraints, including dependence on external resources, frequent servicing

requirements, and vulnerability to harsh outdoor conditions. Many such systems rely on external power supplies, complex mechanical assemblies, or scheduled activation mechanisms, making them unsuitable for decentralized or resource-constrained deployments. Moreover, the absence of adaptive decision-making leads either to unnecessary and frequent cleaning actions, resulting in increased energy consumption and mechanical wear, or to delayed intervention, during which photovoltaic performance remains suboptimal.

Manual cleaning remains the most widely adopted maintenance approach in many regions; however, it is inherently inefficient and prone to human error. The process is labor-intensive, difficult to scale, and often associated with safety risks due to high temperatures, elevated installation locations, and exposure to electrical hazards. From a system management perspective, manual intervention prevents continuous monitoring and timely response to performance degradation, limiting the overall reliability of photovoltaic installations.

Recent advances in automated and dry-cleaning techniques demonstrate the feasibility of reducing water dependency and human involvement. Nevertheless, many existing solutions focus primarily on the mechanical removal of dust**,** while neglecting the information-driven control problem of determining *when* and *why* cleaning should occur. Without real-time sensing, adaptive decision logic, and closed-loop feedback, automated systems remain reactive or schedule-driven rather than truly autonomous.

Consequently, there exists a clear need for an information-centric, autonomous management framework that integrates environmental sensing, performance evaluation, and event-driven control into a unified system. Such a framework must operate with minimal external dependencies, adapt to changing environmental conditions, and maintain photovoltaic performance through intelligent, real-time decision-making. Addressing this gap forms the core problem tackled in this study.

### 1.2. Aim and objectives

The aim of this study is to design and validate an autonomous, information-driven management system for maintaining photovoltaic (PV) performance under dynamic environmental conditions. Rather than focusing solely on mechanical cleaning, the research emphasizes adaptive sensing, real-time decision-making, and closed-loop control for efficient system operation.

The specific objectives of this research are as follows:

- To develop an intelligent autonomous framework that monitors environmental conditions and photovoltaic performance without continuous human supervision.

- To reduce energy losses caused by dust accumulation through event-driven intervention, ensuring timely and efficient performance recovery.
- To implement a predictive reflex control mechanism that activates maintenance actions only when required, avoiding unnecessary actuation and resource consumption.
- To eliminate dependency on manual maintenance by integrating sensing, decision-making, and actuation into a self-operating cyber-physical system.
- To design a low-complexity, scalable, and energy-autonomous solution suitable for decentralized, off-grid, and resource-constrained photovoltaic installations.

### 1.3 Related works

Numerous studies have examined the impact of dust accumulation on the operational efficiency of photovoltaic (PV) systems and proposed various mitigation strategies. Early approaches primarily focused on mechanical or water-based cleaning mechanisms aimed at restoring panel efficiency. While these solutions demonstrated measurable performance improvements, they often relied on predefined schedules, manual supervision, or external resources, limiting their adaptability to dynamic environmental conditions.

Several works have explored low-cost automated cleaning systems for small-scale PV installations. Nkinyam et al. proposed a waterless cleaning system integrating an Arduino-based controller with real-time clock scheduling and a combined air-blower and brush mechanism, achieving an efficiency improvement of 26.23%. Although cost-effective, the system remains schedule-driven and does not incorporate adaptive decision-making based on real-time environmental or performance data. Similarly, Maindad et al. developed a microcontroller-controlled cleaning mechanism employing linear actuators, water pumps, and GSM communication. Despite reducing human involvement, the system remains dependent on water and external infrastructure, which constrains deployment in resource-limited environments.

Other studies have investigated automated cleaning solutions for specific applications, such as solar-powered street lighting systems and multifunctional PV maintenance platforms. Dahoud et al. introduced a GSM-enabled cleaning system for solar lighting installations, while Myyas et al. developed an integrated solution combining cleaning, temperature regulation, and water recycling. Although these systems enhance operational efficiency, they typically involve complex architectures and lack lightweight, autonomous control strategies suitable for decentralized PV deployments.

Robotic and mobile cleaning approaches have also gained attention, particularly for large-scale solar farms. Khadka et al. proposed a rail-guided robotic cleaner with cloud-based monitoring, and Makmee et al. introduced a stretch-sling robotic mechanism to improve positioning accuracy. While effective in controlled environments, such robotic systems often require substantial mechanical

infrastructure, external power sources, and continuous monitoring, limiting their applicability to small or off-grid installations.

In parallel, non-contact dust removal techniques have been explored to reduce mechanical wear and water dependency. Electrodynamic shields, electrostatic cleaning, and surface acoustic wave-based methods have demonstrated potential for dust detachment with low power consumption. However, these approaches generally focus on the physical removal process and provide limited integration with real-time sensing, adaptive control, or system-level performance management.

More recent research has begun to incorporate sensing and classification mechanisms to trigger cleaning actions. For instance, vision-based and color-analysis techniques have been used to estimate dust levels and initiate cleaning. While these methods represent a step toward automation, they often remain reactive and lack closed-loop control architectures capable of continuously managing performance under uncertainty.

In summary, existing research predominantly addresses the mechanical and physical aspects of dust removal, with limited emphasis on information-driven, autonomous system management. The integration of real-time sensing, adaptive decision logic, and energy-autonomous operation into a unified cyber-physical framework remains an open challenge. The present work addresses this gap by proposing a lightweight, event-driven Dust Detection and Protection System (DDPS) that emphasizes autonomous performance management rather than scheduled or manually supervised cleaning.

## 2. System Overview

### 2.1. Hardware Architecture

The hardware architecture of the proposed Dust Detection and Protection System (DDPS) is designed to enable autonomous, real-time decision-making with low complexity and high energy efficiency. The system integrates environmental sensing, embedded control, and actuation into a compact cyber-physical configuration suitable for small- and medium-scale photovoltaic (PV) installations.

Ambient light conditions are monitored using two light-dependent resistor (LDR) sensors, which improve measurement stability and reduce the influence of transient disturbances. Sensor data are processed by an embedded microcontroller to determine the system's operational state. Two potentiometers are used to adjust motor power limits and sensor sensitivity, allowing adaptive operation under varying environmental conditions. A calibration push button enables on-site system tuning by establishing reference light levels.

The control unit is implemented using an Arduino Uno microcontroller, while actuation is performed via a brushless motor controlled through an electronic speed controller (ESC). System feedback is provided through basic visual and audible

indicators. Power is supplied by a 12 V (3S, 11.1 V) battery, enabling standalone and energy-autonomous operation. The overall electrical architecture and signal flow of the DDPS are illustrated in Fig. 2.

**Key hardware components include:** Arduino Uno, LDR sensors, potentiometers, brushless motor with ESC, calibration button, LED, buzzer, battery, and ON/OFF control switch.

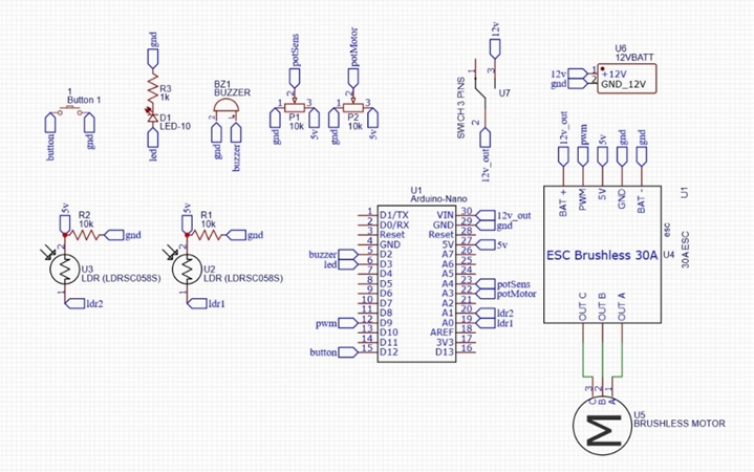



**Fig. 2.** Electrical schematic and hardware architecture of the proposed DDPS.

## 2.2. Software and Algorithm

The Dust Detection and Protection System (DDPS) operates based on an autonomous control algorithm implemented on an embedded microcontroller. The system continuously monitors ambient light conditions using LDR sensors to detect potential performance degradation caused by dust accumulation on the photovoltaic surface.

Sensor readings are compared against calibrated threshold values representing normal operating conditions. When the measured light intensity falls below the defined threshold, the control algorithm activates the cleaning mechanism by starting the motor. Once the light level returns to the normal range, the motor is automatically deactivated. This event-driven logic ensures that cleaning actions occur only when required.

The operational sequence includes system initialization, continuous sensor monitoring, conditional motor activation, and automatic return to standby mode. All processes are executed autonomously without the need for continuous user

intervention. The sensor configuration used for ambient light monitoring is shown in Fig. 3.

**Fig. 3.** LDR sensors

### 2.3. PV module

The photovoltaic (PV) module serves as both the energy generation unit and the physical surface subject to dust accumulation in the proposed system. Variations in incident solar radiation caused by surface soiling directly affect the electrical output of the module, making it a suitable reference element for performance monitoring. In the DDPS framework, changes in illumination and resulting power degradation are used as indirect indicators of dust accumulation, enabling autonomous detection without additional sensing infrastructure. The PV module also supplies electrical energy to the system, supporting energy-autonomous operation.

### 2.4. Control unit (Arduino Uno)

The control unit of the proposed system is implemented using an Arduino Uno microcontroller, which serves as the central processing and decision-making component of the DDPS. The microcontroller acquires sensor data, executes the control algorithm, and generates actuation commands for the cleaning mechanism. The Arduino Uno was selected due to its low power consumption, ease of integration, and suitability for embedded autonomous control applications. Its use simplifies system implementation while ensuring reliable real-time operation in resource-constrained environments.

### 2.5. Cleaning unit

This unit is responsible for removing dust and debris off the surface of the solar panel. The requirement entails cleaning an entire row of solar panels in a PV array

efficiently utilizing low energy without water. This unit performs its cleaning operation at all times. The cleaning method finally chosen contains both a motor and fan. **Fig.4** shows the system responsible for removing dust and debris off the surface of the solar panel. The requirement entails effectively cleaning an entire row of solar panels in a PV array using solar energy without water.

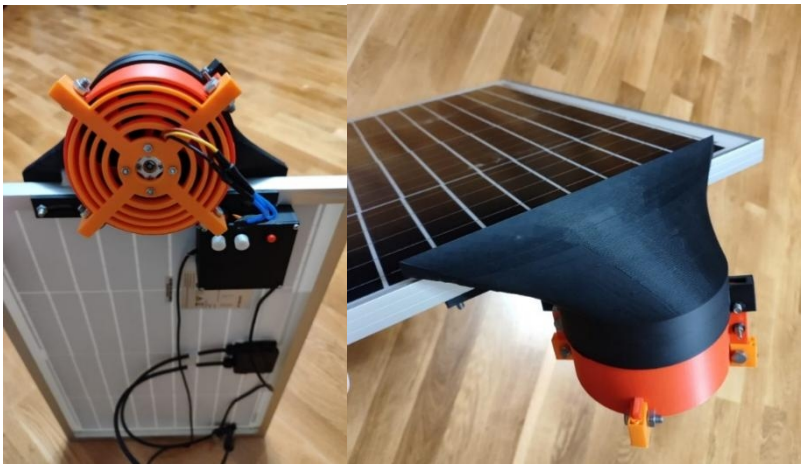


**Fig. 4.** Prototype of cleaning system for solar panels

In this system, the fans serve to generate directed airflows during the cleaning process, dislodging loose sand and dust particles from the surface of the solar panel. The need for solar panel cleaning is highly site-specific and depends on the installation environment, requiring tailored decisions for each photovoltaic system. Several key factors must be considered when selecting appropriate cleaning components, including:

- Annual climatic conditions at the site, such as rainfall, humidity, dew formation, wind patterns, and ambient temperature
- The type and characteristics of surface soiling
- Anthropogenic pollution sources, where high traffic density, urban environments, or industrial activities can significantly increase contamination levels

## 3. Experiment Results

### 3.1. Results

The efficiency of the automatic solar panel cleaning system was assessed under two conditions by comparing the electrical performance of the panels before and after cleaning using a multimeter. Measurements of voltage, current, and power were conducted on both dusty and cleaned panels over a daytime period between 15:00 and

16:30. All measurements were performed with the solar panel fixed at an inclination angle of 20 degrees to ensure consistent irradiation conditions.

### 3.1.1. Measurements on the dusty solar panel

Current and voltage measurements were carried out on the dusty solar panel to determine the voltage and current of the solar panel, applying 120 g of dust on the surface of the panel and measurement are taken at 30 min interval to have several readings.

**Table 1** shows measurements for a dusty solar panel, we observe a general downward trend in voltage, current, and power over time, with each parameter gradually decreasing from 15:00 to 16:30 p.m. The voltage starts at 22.6 V and falls to 21.2 V, while the current drops from 0.20 A to 0.13 A, leading to a decrease in power from 4.52 W to 2.756 W. This decrease corresponds to the reduction in sunlight intensity, which provides less energy for the panel to convert into electrical power.

**Table 1.** Measurements on the dusty solar panel.

| Measurement time | Voltage (V) | Current (A) | Power (W) |
|---|---|---|---|
| 15:00 | 22.6 | 0.20 | 4.52 |
| 15:30 | 21.2 | 0.13 | 2.756 |
| 16:00 | 20.0 | 0.07 | 1.4 |
| 16:30 | 19.1 | 0.01 | 1.91 |

### 3.1.2. Measurements on the clean solar panel

**Table 2** reflects measurements from a clean solar panel, showing a higher overall performance compared to the dusty panel: The voltage starts at 23.1 V at 15:00 and falls to 21.8 V at 15:30, while the current drops from 0.29 A to 0.22 A, leading to a decrease in power from 6.699 W to 4.796 W. The trends mirror those of the dusty panel but at significantly higher values. The absence of dust allows the panel to harness more solar energy, leading to enhanced performance. The current values show a significant increase, reflecting better electron movement within the photovoltaic cells when sunlight is unblocked.

**Table 2.** Measurements on a clean solar panel.

| Measurement time | Voltage (V) | Current (A) | Power (W) |
|---|---|---|---|
| 15:00 | 23.1 | 0.29 | 6.699 |
| 15:30 | 21.8 | 0.22 | 4.796 |
| 16:00 | 20.5 | 0.15 | 3.075 |

| | | | |
|---|---|---|---|
| 16:30 | 19.5 | 0.10 | 1.950 |

**Table 3** illustrates the average increase in output power between the dusty and clean panels. The results showed an increase in the output power of the clean solar panels using our prototype. We calculate the average increase in output power using **Equation (1)**.

$$Q = \frac{c-d}{d} \times 100\% \quad (1)$$

Where Q = Average increase of output power, c = daily power of cleaned solar panel, d = daily power of dirty solar panel.

**Table 3.** Measurements of the average increase of output power.

| **Measurement time** | **Power from dirty panel (W)** | **Power from Clean panel (W)** | **Average increase of output power (Q)** |
|---|---|---|---|
| 15:00 | 4.52 | 6.699 | |
| 15:30 | 2.756 | 4.796 | |
| 16:00 | 1.4 | 3.075 | |
| 16:30 | 1.91 | 1.950 | |
| **daily panel power** | 10.586 | 16.52 | 56.055% |

### 3.2. Discussion

The experimental results confirm that dust accumulation has a substantial impact on the electrical performance of photovoltaic modules. Comparative measurements obtained under dusty and clean conditions show a clear reduction in voltage, current, and output power due to surface contamination.

The clean photovoltaic panel consistently delivers higher output power compared to the dusty panel. The observed average power improvement of 56.055% after cleaning highlights the severity of dust-induced performance degradation and confirms the effectiveness of the proposed Dust Detection and Protection System (DDPS).

More importantly, the results validate the event-driven control strategy employed by the DDPS, in which cleaning actions are initiated based on detected performance degradation rather than fixed schedules or manual intervention. This demonstrates that the system functions as an autonomous performance management solution rather than merely a mechanical cleaning device.

From a system-level perspective, the experimental findings indicate that lightweight

embedded control combined with simple sensing mechanisms can reliably support autonomous operation under varying environmental conditions, making the proposed approach suitable for small-scale photovoltaic installations.

In addition to performance recovery, the DDPS exhibits several system-level advantages relevant to autonomous photovoltaic operation. The waterless and mechanically simple design eliminates dependence on external water resources and reduces mechanical wear. A brushless or minimal-contact cleaning approach enhances system durability, while the predictive, reflex-based control strategy optimizes cleaning frequency and minimizes unnecessary energy consumption. Furthermore, the low-complexity and energy-autonomous configuration enables deployment in small-scale, rural, off-grid, and infrastructure-limited environments.

### 3.3. Anticipated maintenance requirements and technical considerations

The proposed waterless solar panel cleaning system requires periodic maintenance to ensure sustained operational efficiency. Prolonged mechanical operation can also contribute to wear in moving components, potentially decreasing cleaning effectiveness over time. To mitigate premature component degradation, routine lubrication of mechanical elements such as DC motor shafts is advised. Moreover, electronic components may be prone to calibration drift or connectivity issues. Implementing scheduled system diagnostics can help detect and resolve such faults early, thereby minimizing the risk of unexpected system interruptions.

### 3.4. System limitations and operational challenges

In regions experiencing extreme weather events, such as heavy snowfall or elevated humidity levels, the air blower mechanism may be less effective. In such environments, supplementary solutions may be required to maintain cleaning efficiency. While the system demonstrates strong performance in removing dry dust, its effectiveness may be reduced when encountering sticky or oily deposits commonly found in industrial zones or coastal regions. Furthermore, although the current configuration is well suited for small-scale photovoltaic installations, expanding the system to larger PV arrays may necessitate structural reinforcement and increased power capacity to ensure uniform cleaning coverage.

## 4. Conclusion

This paper presented an intelligent Dust Detection and Protection System (DDPS) as an information-driven autonomous control solution for managing photovoltaic system performance under dynamic environmental conditions. By integrating sensor-based monitoring with a predictive reflex control architecture, the proposed system enables

real-time, event-triggered decision-making without reliance on fixed schedules or human supervision.

Experimental results demonstrate that the DDPS effectively detects performance degradation and initiates corrective action, achieving a measurable restoration of photovoltaic output following dust accumulation. These findings validate the effectiveness of reflex-based autonomous control in maintaining system performance while minimizing unnecessary actuation, energy consumption, and mechanical wear. The closed-loop nature of the system ensures adaptive behavior in response to environmental uncertainty, a critical requirement for modern cyber-physical energy systems.

A key advantage of the proposed architecture is its energy-autonomous operation, as all system components draw power directly from the photovoltaic panel itself. This characteristic enhances resilience and enables deployment in off-grid, remote, or high-risk environments where external power supply and maintenance access are limited. Furthermore, the system's modular and low-complexity design supports scalability and integration into small and medium-scale photovoltaic installations.

From an information technology perspective, the DDPS exemplifies how embedded sensing, lightweight control logic, and autonomous decision-making can be combined to manage complex physical processes efficiently. While this study focuses on photovoltaic performance management, the underlying architecture is broadly applicable to other domains requiring autonomous monitoring and control under uncertainty, such as smart infrastructure, environmental monitoring, and distributed energy systems.

Future work will extend the DDPS framework by incorporating distributed communication technologies such as LoRaWAN and Zigbee, enabling cooperative management across multiple photovoltaic units. Additionally, integrating predictive models based on weather data and machine learning techniques for dust accumulation forecasting could further enhance system intelligence and long-term performance optimization.

## References

1. Adekanbi, M. L., Alaba, E. S., John, T. J., Tundealao, T. D., & Banji, T. I. (2024). Soiling loss in solar systems: A review of its effect on solar energy efficiency and mitigation techniques. Cleaner Energy Systems, 7, 100094.
2. Al Dahoud, A., Fezari, M., & Al Dahoud, A. (2021, November). Automatic solar panel cleaning system Design. In 2021 29th Telecommunications Forum (TELFOR) (pp. 1–4). IEEE.
3. Alagoz, S., & Apak, Y. (2020). Removal of soiling materials from solar panel surfaces by applying surface acoustic waves. Journal of Cleaner Production, 253, 119992.

4. Al-Housani, M., Bicer, Y., & Koç, M. (2019). Experimental investigations on PV cleaning of large-scale solar power plants in desert climates: Comparison of cleaning techniques for drone retrofitting. Energy Conversion and Management, 185, 800-815.
5. Amin, A., Wang, X., Alroichdi, A., & Ibrahim, A. (2023). Designing and Manufacturing a Robot for Dry-Cleaning PV Solar Panels. International Journal of Energy Research, 2023(1), 7231554.
6. ASEU. (2023, March 4). Ukrainians shift to renewable power for energy security amid war.
7. Badalyan, N. K. (2024). Analysis of Soiling Measurement Systems and Cleaning Methods of Solar Photovoltaic Modules. Proceeding of NPUA: Electrical Engineering, Energetics, 1, 74-83.
8. Bessa, J. G., Micheli, L., Almonacid, F., & Fernández, E. F. (2021). Monitoring photovoltaic soiling: assessment, challenges, and perspectives of current and potential strategies. iScience, 24(3).
9. Boulaid, M., Oaddi, R., Tihane, A., Elfanaoui, A., & Ihlal, A. (2020). Photovoltaic system energy yield monitoring. Materials Today: Proceedings, 24(Part 1), 81–84.
10. Burke, M., Greenough, R., Jensen, D., & Voss, E. (2016). Project SPACE: Solar Panel Automated Cleaning Environment. Mechanical Engineering Senior Theses, 62.
11. Crawford, M. (2022). Water-Free Cleaning of Solar Panels Saves Millions of Gallons of Water. ASME.
12. Ghosh, A. (2020). Soiling losses: A barrier for India's energy security dependency from photovoltaic power. Challenges, 11(1), 9.
13. Hooshyar, P., Moghadasi, H., Moosavi, S. A., Moosavi, A., & Nouri Borujerdi, A. (2025). Recent progress of soiling impact on solar panels and its mitigation strategies: A review. Applied Energy, 379, 124979.
14. Kalaji, A. A., Abbas, A. S., & Magdy, M. (2022). A review of automatic cleaning systems for photovoltaic solar panels. International Advanced Research Journal in Science, Engineering and Technology, 9(1).
15. Kawamoto, H. (2019). Electrostatic cleaning equipment for dust removal from soiled solar panels. Journal of Electrostatics, 98, 11–16.
16. Khadka, N., Bista, A., Adhikari, B., Shrestha, A., & Bista, D. (2020, March). Smart solar photovoltaic panel cleaning system. In IOP Conference Series: Earth and Environmental Science (Vol. 463, No. 1, p. 012121). IOP Publishing.
17. Li, D., King, M., Dooner, M., Guo, S., & Wang, J. (2021). Study on the cleaning and cooling of solar photovoltaic panels using compressed airflow. Solar Energy, 221, 433–444.
18. Maindad, N., Gadhave, A., Satpute, S., & Nanda, B. (2020, April). Automatic solar panel cleaning system. In 2nd International Conference on Communication & Information Processing (ICCIP).
19. Makani, J. M., Patel, V. B., Sharma, A. J., Patel, M. M., & Patel, H. A. (1954). Design of waterless solar panel cleaning system. Int. Res. J. Eng. Technol, 1958.
20. Makmee, T., & Pengwang, E. (2021, April). Design and experiment of four-side stretch sling-based solar cleaning robot. In 2021 7th International Conference on Engineering, Applied Sciences and Technology (ICEAST) (pp. 182–189). IEEE.
21. Mekonnen, M. M., Gerbens-Leenes, P. W., & Hoekstra, A. Y. (2015). The consumptive water footprint of electricity and heat: a global assessment. Environmental Science: Water Research & Technology, 1, 285–297.
22. Min, R. (2024, February 13). Palestinians rely on solar panels amid electricity cuts to talk to friends and family. Euronews.

23. Myyas, R. E. N., Al-Dabbasa, M., Tostado-Véliz, M., & Jurado, F. (2022). A novel solar panel cleaning mechanism to improve performance and harvesting rainwater. Solar Energy, 237, 19–28.
24. Nahar Myyas, R., Al-Dabbasa, M., Tostado-Véliz, M., & Jurado, F. (2022). A novel solar panel cleaning mechanism to improve performance and harvesting rainwater. Solar Energy, 237, 19–28.
25. Nkinyam, C. M., Ujah, C. O., Nnakwo, K. C., Ezeudu, O., Kallon, D. V. V., Ike-Eze, I., & Ezema, C. (2025). Design and implementation of a waterless solar panel cleaning system. Unconventional Resources, 5, 100131.
26. Otter, P., Sattler, W., Grischek, T., Jaskolski, M., Mey, E., Ulmer, N., Grossmann, P., Matthias, F., Malakar, P., Goldmaier, A., Benz, F., & Ndumwa, C. (2020). Economic evaluation of water supply systems operated with solar-driven electro-chlorination in rural regions in Nepal, Egypt and Tanzania. Water Research, 187, 116384.
27. Patel, S., Veerasamy, V. S., John, J. S., & Orlov, A. (2023). A comprehensive review on dust removal using electrodynamic shield: Mechanism, influencing factors, performance, and progress. Renewable and Sustainable Energy Reviews, 183, 113471.
28. Topaloglou, L., & Nikolaidou, K. (2025). Exploring Climate, Energy, and Environmental Justice in Climate Change, SDGs, and Transition Contexts.
29. White, W. J., O'Connor, A. C., & Rowe, B. R. (2004). Economic Impact of Inadequate Infrastructure for Supply Chain Integration. National Institute of Standards and Technology Planning Report 04-2. Research Triangle Park: RTI International.

# Advanced Modelling of Chloride Penetration in Concrete Considering Time- and Space-Dependent Diffusivity

Enrico Zacchei [1] [0000-0003-2537-0860] and Md. Munir Hayet Khan [2] [0000-0002-0947-5008]

[1] Department of Mechanical Engineering, Polytechnic School of the University of Salamanca (USAL), Avila, Spain
enricozacchei@usal.es
[2] Faculty of Engineering & Quantity Surveying (FEQS), INTI International University, Persiaran Perdana, Putra Nilai 71800 Nilai, Negeri Sembilan Malaysia

**Abstract.** For reinforced concrete structures, corrosion of reinforcing steel bars induced by chloride ions represents a recurring durability risk. Chloride ion diffusion is commonly described by Fick's laws, which are strictly applicable only when the diffusivity is assumed to be constant in both time and space. Under this assumption, the diffusion process and the associated exposure scenario can be greatly simplified. In this paper, a more comprehensive, non-constant, multifactorial diffusivity model is proposed, accounting for, among others, the effects of temperature, deformation, and damage. Given that the diffusivity varies in time and space, random numerical solutions are required. For this purpose, the Runge-Kutta method is implemented for both 1D and 2D models, and the results are plotted using Hermite polynomials. The boundary conditions relevant to this application are calibrated to solve the governing partial differential equations under randomly fluctuating diffusivity. The results demonstrate that traditional approaches significantly underestimate chloride concentrations at certain depths, particularly for long-term exposure scenarios.

**Keywords:** Diffusion model; corrosion; advanced analyses; structural engineering.

## 1 Introduction

One of the most relevant deterioration mechanisms of reinforced concrete (RC) structures is the corrosion of reinforcing steel bars caused by chloride ion diffusion. Chloride ions penetrate the concrete cementitious matrix from the external environment.

In general, structures located in aggressive environments, such as industrial areas and coastal zones up to about 4.0 km from the sea, are highly susceptible to this type of attack [10]. Several studies on real structures (e.g., bridges, offshore structures, etc.) have shown that proximity to the sea has a significant impact [11].

Once corrosion begins, it reduces the cross-sectional area of the steel bars, decreasing their resistance and ductility, with consequences for the structure's stability, load-bearing capacity, and public safety [3, 12, 13].

Figure 1 shows a real photo of two RC beams subjected to corrosions of the steel bars and concrete [1]. The purpose of these photos is only to help illustrate the risk associated with the occurrence of this phenomenon, which must be avoid in order to maintain and "build resilient infrastructure" in accordance with the 2030 Agenda for sustainable development [16].

**Fig. 1.** Real photo of the RC beams subjected to corrosions.

The chloride ions diffusion in concrete is a relatively simple process, since it is a non-linear process depending on the time and several other intrinsically phenomena, such as the capillary absorption and permeation of the water contenting chloride and other ions. The diffusion process is described by Fick's laws, where the diffusion coefficient (or diffusivity) is the key parameter to estimate the trend of the chloride concentration in concrete [17, 18].

In fact, credible papers put the attention on the correct estimation of the diffusivity in order to account for chemical and mechanical interactions, which can be distinguished in four different types as self-diffusivity, free diffusivity, diffusivity in saturated capillary pores, solid diffusivity [7].

The diffusivity is largely used as a constant value in a determinist way, usually dependent only on the water/cement (w/c) ratio, which can be considered as a reference and standard value. However, in this way, some important factors are not considered as the binding chloride, temperature, humidity, concrete aging, etc. (treated in this study). In [14, 19] some empirical and analytical models for predicting the chloride diffusivity are listed, where these factors are considered but, in many cases, separately to each

[1] These photos refer to the fire station building built in 1984 and located in the city of Penela, Portugal. The first author of this paper worked on this executive project during the years 2019-2020 to evaluate its stability and resistance for seismic retrofitting in the presence of a new construction built above [5].

other. Therefore, considering a unique formula where more factors are accounted for could be an innovative aspect to be studied.

These models make the diffusivity non-constant and highly non-linear. Therefore, the use of a constant diffusivity could be insufficient to represent more accurate and realistic chloride concentrations in concrete.

The correlation between diffusivity and temperature could be relevant when considering climate change, as highlighted by the intergovernmental panel on climate change (IPCC) [15], which may require the re-evaluation of many existing structures considering this factor as studied in [3].

Also, the representation in only one-direction (1D) can be very simplified not representing the real situations. For this, the focus of this paper is to quantify a more complete diffusivity, in order to include different factors that mainly account for the material characteristics and environmental variations in 1D and 2D models. These analyses are only possible by developing advanced methods as Runge-Kutta (RK) and Hermite polynomials (HPs). Therefore, the main purpose of this study is definitively to overpass the use of the constant diffusivity by considering external and internal factors.

## 2 Materials and Methods

Consider a structure near a coastal area that, for years, has been exposed to extreme conditions and a corrosion phenomenon. A chemical element that plays an important role for these processes is the chloride ion.

The total concentration of chloride ions, C, in direction i (x, y, z) and time, t, in a homogeneous and saturated concrete structural member is governed by the partial differential equation (PDE) called second Fick's law:

$$\frac{\partial \mathrm{C(i,t)}}{\partial \mathrm{t}} = \mathrm{D}\frac{\partial^2 \mathrm{C(i,t)}}{\partial \mathrm{i}^2} \tag{1}$$

where D is the diffusivity that can be expressed by using a relation accounting for many multi-factors as [1, 4]:

$$\mathrm{D} = \mathrm{f_0}(w/c)\mathrm{f_1}(\mathrm{C_b})\ \mathrm{f_2}(\mathrm{T})\ \mathrm{f_3}(\mathrm{t})\ \mathrm{f_4}(\varepsilon, \mathrm{d})\ \mathrm{f_5}(\mathrm{h}) \tag{2},$$

which each factor, $f_0$, $f_1$, $f_2$, $f_3$, $f_4$, and $f_5$ is correlated to the water-cement (w/c) ratio, binding chloride $C_b$, internal temperature T, time (or age) t, elasto-plastic axial deformation ε, concrete damage d, and internal humidity h, respectively.

Some factors are more common to be used as for instance T and h, whereas other ones are more difficult to be estimated as ε and d. In literature (among others [7, 8]) several analytical and experimental relations can be found, which are valid under some hypotheses to be known to control the model and to quantify the outputs. In this study, an

interesting and non-common aspect regards the approach in adopting a unique relation accounting for many factors simultaneously.

The total C value in Eq. (1) is divided in two parts: the free chloride dissolved in pore solution, $C_f$, and the mentioned $C_b$, where their correlation is defined by the well-known non-linear Langmuir isotherm [2] mainly depending on the evaporable water content, and characteristics of the material composition. In this sense, only $C_f$ participates to the mechanical diffusion process being "free" to move and for this reason it is treated in this paper, whereas $C_b$ is entrapped (i.e., "binding") between other ions hence it could be studied by chemical analyses.

Equation (1) can be solved analytically under specific conditions and considering D as constant parameter, or numerically in order to consider non-constant D. In particular, for $D \neq 1$, it is possible to use the explicit Runge-Kutta (RK) method to find the solutions and Hermite polynomials (HPs) to plot them [3].

The ordinary HPs in x (i.e., $H_n(x)$) and t (i.e., $H_n(t)$), described by the numerical series, generating functions, and identities functions, are, respectively:

$$\begin{cases} H_n(x) = n! \sum_{r=0}^{\frac{n}{2}} \frac{(2x)^{n-2r}(-1)^r}{(n-2r)!r!}, & \text{in x direction} \\ H_n(t) = n! \sum_{r=0}^{\frac{n}{2}} \frac{(2t)^{n-2r}(-1)^r}{(n-2r)!r!}, & \text{in time t} \end{cases} \tag{3}$$

$$\begin{cases} \sum_{n=0}^{\infty} \frac{p^n}{n!} H_n(x) = e^{2xp+p^2}, & \text{in x direction} \\ \sum_{n=0}^{\infty} \frac{p^n}{n!} H_n(t) = e^{2tp+p^2}, & \text{in time t} \end{cases} \tag{4}$$

$$\begin{cases} H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right), & \text{in x direction} \\ H_n(t) = (-1)^n e^{t^2} \frac{d^n}{dt^n}\left(e^{-t^2}\right), & \text{in time t} \end{cases} \tag{5}$$

where the indexes n, r, and p provide the order of the formulas and define their values (for more details see [6]). These HPs satisfy the orthogonality relation, i.e., $H_n(x)$ and other HPs of degree j, $H_j(x)$, are orthogonal in $L^2_\omega(\Lambda)$, where $\Lambda = (-\infty, +\infty)$ regards the weight function defined as $\omega(t) = e^{-x^2}$, and they satisfy the following formula:

$$\int_{-\infty}^{+\infty} \left[H_n(x) H_j(x) \omega(x)\right] dx = 2^n n! \sqrt{\pi} \delta_{n,j} \tag{6}$$

where $\delta_{n,j}$ is the Kronecker delta function under n and j subscripts. For instance, for $H_n(t)$ in t-domain, Eq. (6) assumes a similar form with x = t, whereas Eq. (6) is equal to zero when $n \neq j$.

In this study, to solve Eq. (1) analytically by considering a non-constant D, the initial and boundary conditions (i.c. and b.c., respectively) must be defined as:

$$\begin{aligned} &\text{i.c.:} \quad C(x_m, 0) = C_0, \quad \text{in } \Omega \\ &\text{b.c.:} \quad C(0, t_m) = C_s, \quad \text{at } \partial\Omega_1 \\ &\text{b.c.:} \quad C(q\, x_m, t_m) = C_0, \quad \text{at } \partial\Omega_2 \end{aligned} \tag{7}$$

where $x_m$ and $t_m$ are two variables related to the position and time of the chloride concentration, respectively, which must be pre-defined to not generate the arbitrary constants that represent the degrees of freedom during a mathematical integrating operation; for this Eq. (7) refers to a finite medium (see Fig. 2 (left)). The domains $\Omega$, $\Omega_1$, $\Omega_2$ represent the concrete surface and its boundaries in a more general way.

The parameters $C_0$ and $C_s$ are the inner chloride concentration already present in a concrete member during its manufacturing and the surface chloride concentration transported by wind, respectively.

Note that Eqs. (3)-(7) refer to a 1-dimensional (1D) model in a x direction, however it is possible to extend this model in a 2D by adding the polynomial $H_n(y)$ in y-direction, which can be estimated in analogue way to $H_n(x)$. Due to this similarity for brevity only some formulas are shown.

The parameter q in Eq. (7) quantifies the goodness of the approximated analysis with respect the "exact" one. The exact analysis can be obtained by Eq. (1) under a constant D, since in this way it is possible to find the closed-form analytical solution (i.e., the traditional common solution) of the semi-infinite medium shown in Fig. 2 (right) and used in [17]. The convergence between both analyses has been obtained by several interactions, and it was considered reached when the mean relative error in terms of chloride concentrations provides the minimum value (or an error < 5.0%, usually accepted in structural engineering). Finally, a liner q-D function is plotted providing thus an acceptable value of $q \approx 12.50$ (for more details see [1, 4]).

This q-parameter is also the key to pass from a traditional analysis with a constant diffusivity in 1D semi-infinite medium to an advanced analysis with non-constant diffusivity in 2D finite medium as shown in Fig. 2. Finding a defined q value appears useful for this purpose.

In Fig. 2 it is represented a 1D constant diffusivity (dashed line in right side) and a 2D non-constant and random diffusivity (left side). All parameters have been already explained.

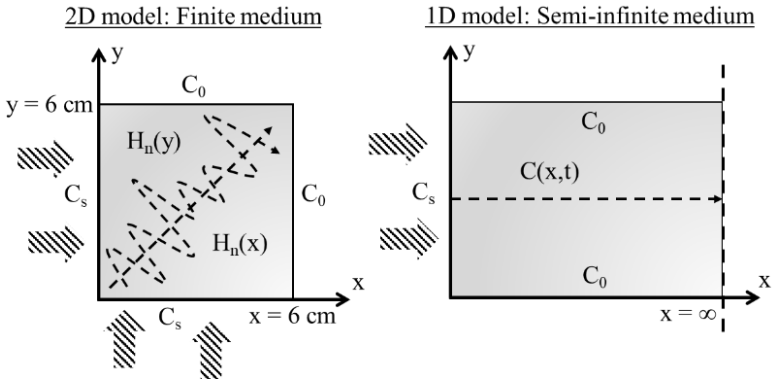



**Fig. 2.** General scheme of the chloride diffusion by analytical 1D and numerical 2D model.

Table 1 lists the main parameters used to carry out this study [7, 8].

**Table 1.** Used parameters for the analyses.

| Parameter | Value |
|---|---|
| Inner chloride concentration, $C_0$ | 0.50 kg/m$^3$ |
| Surface chloride concentration, $C_s$ | 1.15 kg/m$^3$ |
| Reference factor, $f_0$ (for w/c = 0.50) | 4.35 cm$^2$/year |
| Binding factor, $f_1(C_b)$ | 0.66 |
| Temperature factor, $f_2$ (for T = 20.0 ºC) | 0.84 |
| Age factor, $f_3(t)$ | 1.28 |
| Deformation and damage factor, $f_4(\varepsilon, d)$ | 1.0 – 8.92 |
| Humidity factor, $f_5$ (for h = 60.0%) | 0.13 |
| Diffusivity, D | 0.41 – 3.66 cm$^2$/year (Eq. (2)) |

Note that D parameter ranges between two limits due to the variability of the $f_4$ factor shown in [19], whereas the $f_0$ factor provides its physical meaning.

In the sequel the analyses and results are plotted and discussed.

## 3 Analyses and Results

Numerical analyses under the mentioned data are here developed and results are plotted in terms of chloride concentrations in a concrete element. Figure 3 shows the chloride concentration by exact analytical and numerical solutions for a constant D = 0.41 cm$^2$/year. Results during a period of t = 3.0 - 57.0 years and concrete depth x = 0 - 6.0 cm are shown.

Figure 3a) shows non-calibrated results (i.e., q = 1.0) whereas Fig. 3b) shows the calibrated results (q = 12.50) indicating a very good agreement. This comparison allows to understand well the weight of the q variation. It is also possible to see that the boundary conditions at x = 0 play an important role on the calibration since in the non-calibrated solution $C \neq C_s$, which is inconsistent to Eq. (7).

Note that the chloride concentration, obtained by the numerical approach, has been calculated considering the multi-factorial D (Eq. (2)) usually neglected.

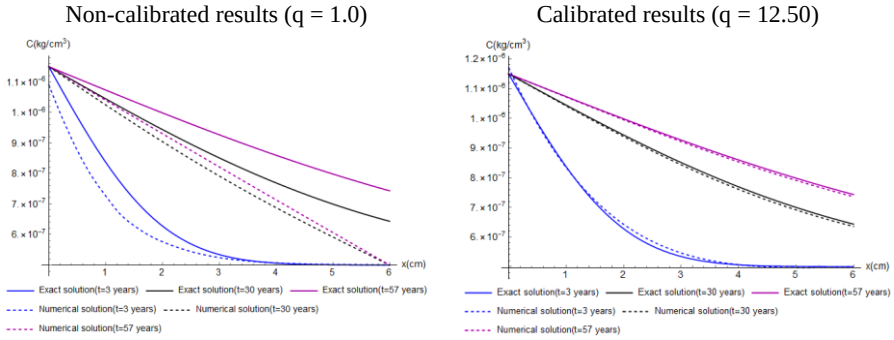



**Fig. 3.** Chloride concentrations by analytical and numerical solutions under a) non-calibrated q, and b) calibrated q value.

A finite number of random combinations in order to perform and to show the analysis has been chosen. Here, 3 combinations have been considered to represent 3 possible different cases by using the classical methods to develop general loadings [9]:

$$D(t) = \frac{D}{\alpha}\left[\sum_{i=1}^{N} \sin(\omega_i t + \phi_i) + \alpha\right] \quad (8)$$

where $\omega_i$ is the circular frequency, $\phi_i$ is the random phase ($\phi_i$ = 0 - 2$\pi$), N and $\alpha$ are partial factors purely correlated to the D form, e.g., N = 10, $\alpha$ =10 allows to generate one or two peaks each year by simulating season peaks (winter and summer).

Figure 4 shows the generated amplitude of the diffusivity (constant and random lines in Fig. 4a)) that produce the chloride concentration curves shown in Fig. 4b). The constant values refer to the values shown in Table 1. The considered points are placed at 3.0 cm from the external surface of the concrete structural member. The period interval for the analysis is 0 - 60.0 years.

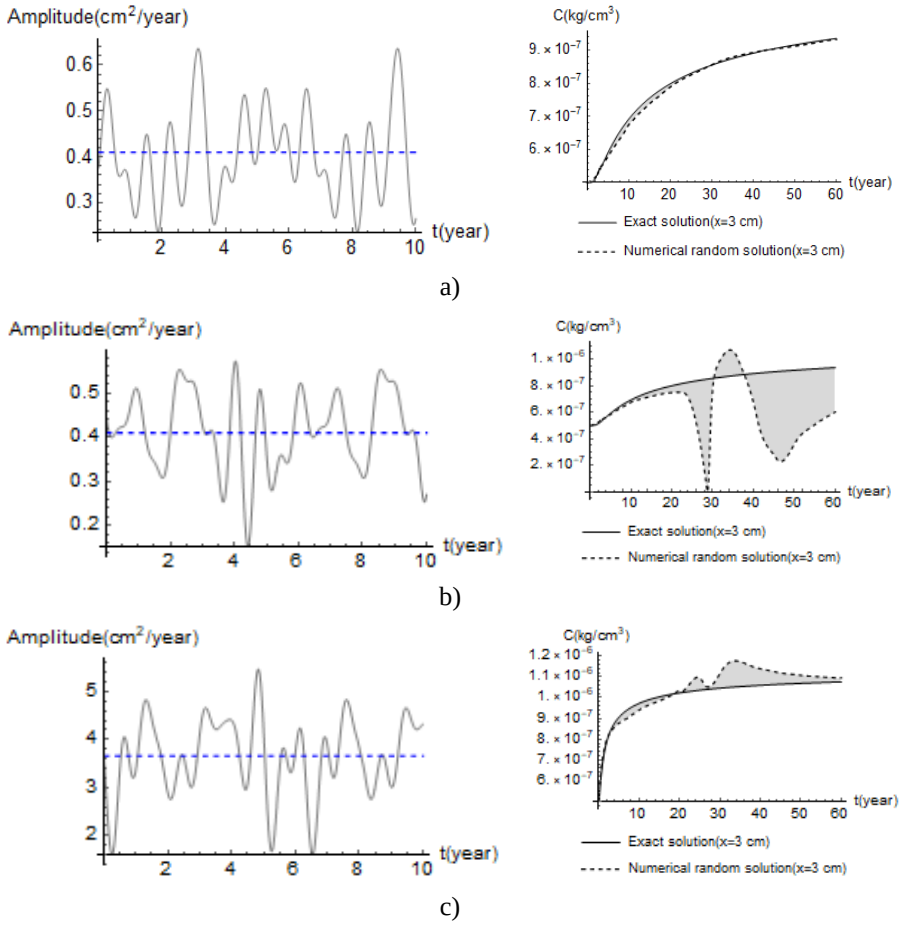



**Fig. 4.** Constant and random D amplitude (left side) and chloride ions concentrations (right side) for a) random 1, b) random 2, and c) random 3.

The chloride concentration by random D can be greater or lesser than the chloride concentration by constant one. The trend follows the principle that C increases with increasing D. When a diffusivity trend is composed by more points placed above the constant D, the chloride concentration by numerical solution is greater than the chloride concentration by exact solution and vice-versa. The solid grey area (in Fig. 4, right) represents the difference of the chloride concentration in concrete between two solutions. In this way, it is possible to quantify the possible over or under-estimation of concentration C.

The chloride concentration is plotted for a defined depth x. Considering that the proposed model is non-linear dynamic, it is possible that for the same depth x, the chloride concentration C decreases or increases in the time t, differently for the constant D, which always increases the concentration C monotonically. With this approach, it is

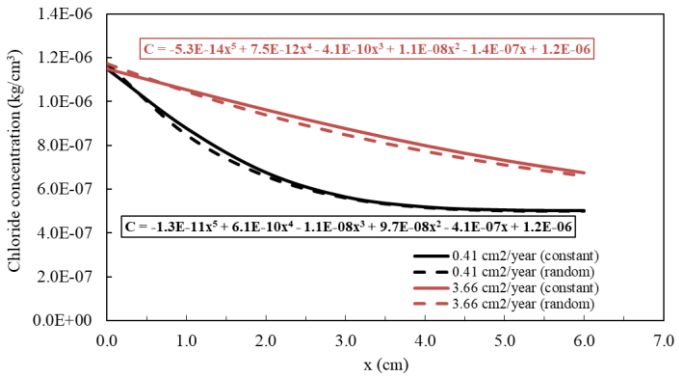
possible to model the random nature of the chloride ions flux in the cementitious matrix simulating real applications.

The values of the chloride concentration in Fig. 4c) are higher than those in Fig. 4a)-b) up to about 1.35. This is due to the different vales of D and the amplification of the $f_4$ factor, which accounts for the elastic deformation like a structure response under external loadings. This is an important aspect worth emphasizing is that chloride ion flux always occurs when structures support service loads; therefore, the effects of stress-strain states along the structure, including deformation and damage, must be considered in chloride penetration prediction models for more accurate assessments.

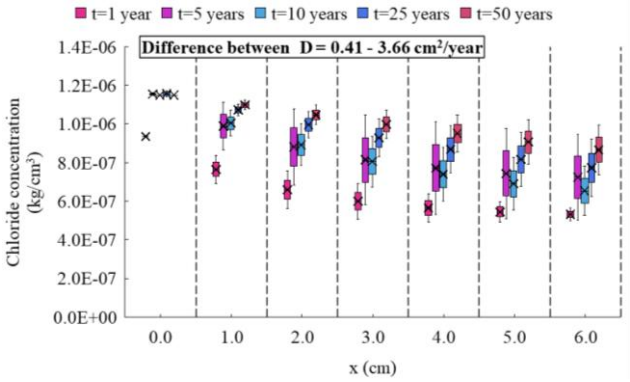
It is possible to note that relevant effects of the random D occur after a certain time (in general, t > 25.0 years). This could indicate that for large periods, the use of constant D bad estimates the scenarios in RC structures.

By Eq. (3), the cubic polynomial approximate solutions in time t, and fifth degree polynomial approximate solution in x direction are obtained. Figure 5a) shows the chloride concentration by analytical (solid lines) and numerical (dashed lines) solution for both limits of D. The approximated polynomials refer to trends similar to random 1 where the discrepancies are minimal, thus the relations in Fig. 5a) could be considered more reliable and useful for practical applications.

Figure 5b) shows the difference between both diffusivity limits for different years. The difference of the values are represented by using a box and whisker chart with a mean value indicated by "×". The upper extreme of the box chart refers to the chloride concentration obtained by D = 3.66 $cm^2$/year, whereas the lower extreme refers to 0.41 $cm^2$/year. In general, the boxes increase by increasing x-depth, however due to the randomness of the analysis this trend is not always verified.




a)




b)

**Fig. 5.** Chloride concentration in terms of a) polynomial trends (t = 4.0 years), and b) general variations.

Finally, as already mentioned, the analysis can be extended from 1D to 2D model with D = 3.66 cm$^2$/year for the period t = 1.0 – 50.0 years. The interpolation for 2D model is carried out before in one direction, x-axis, and then in other direction y-axis (or vice-versa), thus the chloride concentrations inside concrete are plotted.

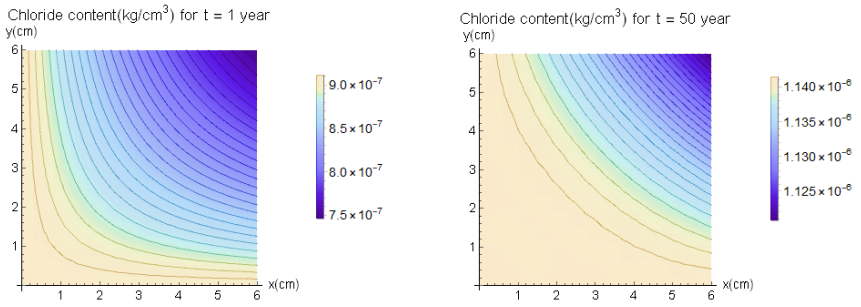



**Fig. 6.** 2D chloride concentrations by numerical random solutions for D = 3.66 cm$^2$/year at 1.0 - 50.0 years.

In analogue way for Fig. 4, for large time the effects of a random D more influence the chloride concentration as visible in Fig. 6. This is consistent to the physical real process, since the chloride concentration is a cumulative process that comes from different directions and not only from one direction are usually modelled. Actually, this is more evident in 2D model. The mean 2D/1D ratio by considering random and analytical model is about 1.25 (i.e., 1.14/0.90). This gap should quantify the poor approximation of the analytical model by using constant diffusivity in only one direction.

## 4 Conclusions

In this paper the chloride ions diffusion in RC members by considering non-constant multi-factorial diffusivity has been studied. For this, numerical solutions in 1D and 2D models have been developed by using explicit Runge-Kutta (RK) method to find the solutions and by Hermite polynomials (HPs) to plot them. Specific boundary conditions have been specified for this issue.

A more complete diffusivity D accounts for the w/c ratio, binding chloride, temperature, concrete age, concrete deformation, damage, and humidity. From literature, it appears that these factors are the most relevant to well quantify the diffusivity thus the chloride concentrations.

By using a non-constant random D, it is shown that the analytical solutions poor estimates the chloride concentration in RC structures for large period. In particular, the impact of the random D happens after a certain time. Due to the dynamism of the proposed model, it is possible that the chloride concentration C, at a certain depth x, decreases in the time t, whereas the total concentration C, in a RC element, increases. Also, the non-linearity of the random diffusivity D could provide low values in 1D model and high values in 2D model. These phenomena show the low reliability of the analytical solution using a constant D for predicting long scenario and instantaneous chloride concentrations.

## Acknowledgements

The first author acknowledges the University of Salamanca (USAL), Spain, for providing a Wolfram Mathematica license. The authors acknowledge their institutions for covering the publication access fees, when applicable, required to download the referenced papers.

## References

1. Zacchei, E., Nogueira, C.G., Calibration of boundary conditions correlated to the diffusivity of chloride ions: An accurate study for random diffusivity, Cement and Concrete Composites 126, 1-14, 2022.
2. Chen, W., Zhang, J., Zhang, J.A, Variable-order time-fractional derivative model for chloride ions sub-diffusion in concrete structures, Fract Calc Appl Anal. 16, 76-92, 2013.
3. Zacchei, E., Hoult, R.D., De Almeida, J.P., Influence of chloride and corrosion attack on the seismic resistance of reinforced concrete walls under alternative climate scenarios, Structural Concrete 26, 1-20, 2025.
4. Zacchei, E., Nogueira, C.G., 2D/3D numerical analyses of corrosion initiation in RC structures accounting fluctuations of chloride ions by external actions, KSCE Journal of Civil Engineering 25, 1-16, 2021.
5. Zacchei, E., Brasil, R., Calculation of Seismic Actions for Structural Analysis - Methodological Approaches and Experiences, Springe Nature, p. 107, 2025.

6. Dattoli, G., Generalized polynomials, operational identities and their applications, J. Comput. Appl. Math. 118, 111-123, 2000.
7. Carrara, P., De Lorenzis, L., Bentz, D.P. Chloride diffusivity in hardened cement paste from microscale analyses and accounting for binding effects, Model. Simul. Mater. Sci. Eng. 24, 1-44, 2016.
8. Bastidas-Arteaga, E., Chateauneuf, A., Sanchez-Silva, M., Bressolette, P., Schoefs, F., A comprehensive probabilistic model of chloride ingress in unsaturated concrete, Eng. Struct. 33, 720-730, 2011.
9. Clough, R.W., Penzien, J., Dynamics of Structures, McGraw-Hill, New York, USA, 3rd ed, p. 752, 2003.
10. Valdes, C., Castañeda, A., Corvo, F., Marrero, R., Montero, R., Atmospheric corrosion study of carbon steel in Havana waterfront zone, Proceedings of the International Conference of Sustainable Production and Use of Cement and Concrete, RILEM 22, 1-10, 2020.
11. Othmen, I., Bonnet, S., Schoefs, F., Statistical investigation of different analysis methods for chloride profiles within a real structure in a marine environment, Ocean Eng. 157, 96–107, 2018.
12. Dizaj, E.A., Madandoust, R., Kashani, M.M., Probabilistic seismic vulnerability analysis of corroded reinforced concrete frames including spatial variability of pitting corrosion, Soil Dyn Earthq Eng. 114, 97-112, 2018.
13. Shekhar, S., Freddi, F., Ghosh, J., Lad, D., Influence of corrosion on failure modes and lifetime seismic vulnerability assessment of low-ductility RC frames, Earthq Eng Struct Dyn. 52, 1-23, 2023.
14. Farahani, A., Taghaddos, H., Shekarchi, M., Prediction of long-term chloride diffusion in sílica fume concrete in a marine environment, Cement Concr. Compos. 59, 10–17, 2015.
15. Intergovernmental Panel on Climate Change (IPCC), Technical report., 2018. Accessed on 02/2026: https://www.ipcc.ch/sr15/graphics/
16. United Nations (UN), Transforming Our World: The 2030 Agenda for Sustainable Development, Technical Report, New York, USA, p. 35. A/RES/70/1. Accessed on 02/2026: https://sdgs.un.org/goals
17. Nogueira, G.C., Yoshio, L., Zacchei, E., Deterministic and probabilistic approaches for corrosion in RC structures: A direct proposed model to total service life predictions, Case Studies in Construction Materials 18, 1-20, 2023.
18. Zacchei, E., Avaliação da difusividade de íons cloretos em estruturas de concreto armado considerando fatores ambientais e decorrentes das ações externas: uma aplicação na estimativa determinística e probabilística do tempo de início de corrosão, PhD Thesis, University of São Paulo State (UNESP), Bauru, Brazil, p. 153, 2020.
19. Fu, C., Jin, X., Ye, H., Jin, N., Theoretical and experimental investigation of loading effects on chloride diffusion in saturated concrete, J. Adv. Concr. Technol. 13, 30-43, 2015.

# Clustering-Based Post-Processing of Experimental Data for Double Glass Units

Enrico Zacchei [1] [0000-0003-2537-0860] and Zhiyuan Wang [2,3] [0009-0004-7282-0363]

[1] Department of Mechanical Engineering, Polytechnic School of the University of Salamanca (USAL), Avila, Spain
enricozacchei@usal.es
[2] China Academy of Building Research, Beijing 100013, China
[3] Department of Civil Engineering, Tongji University, Shanghai 200092, China

**Abstract.** Double glass units (DGUs) consist of glass plates separated by a gas-filled cavity sealed by perimeter spacers. The glass plates resist to applied loads, whereas the enclosed gas mainly contributes to thermal and acoustic insulation in buildings or other structures. Load transfer between the plates occurs through a load-sharing mechanism that is difficult to be quantified, particularly regarding its influence on vertical deflections of the glass. Although this phenomenon has been recently investigated experimentally, a major challenge of laboratory testing is the post-processing of large outputs. To address this issue, this study employs the k-means clustering algorithm as a data-driven tool for experimental data analysis. The objective is to identify optimal clusters that provide accurate approximations and enable direct relationships between elastic stiffness and deflections. The results show that k-means clustering produces reliable stochastic approximations, allowing the generation of deflection maps for both the upper and lower glass plates. Compared with traditional manual or analytical post-processing methods, the proposed approach offers an efficient and robust alternative for estimating stiffness values and glass deflections.

**Keywords:** k-means; DGU; mechanical deflections; structural analysis.

## 1 Introduction

With the progressive development and widespread adoption of the artificial intelligence, numerous algorithms have been employed to model input-output relationships with high precision. This approach significantly reduces human workload and enhances the efficiency of data processing and analysis.

The k-means algorithm is used for processing several parameters and it appears particularly effective for analysing datasets with multiple variables and dimensions [12]. It becomes especially valuable when dealing with "big data" issue, where conventional methods often fall short. As noted in [10] the greater the number of predictor variables, the harder it is to interpret and isolate each predictor variable effect; also, clus-

tering analysis can help identify data patterns, merge similar components, and prepare the groundwork for predictive model construction. Due to its simplicity, efficiency, and versatility, the k-means algorithm has been widely applied across various disciplines beyond structural and mechanical engineering.

Therefore, this approach has provided the foundational inspiration for the present work in developing the k-means algorithm also for double glass units (DGUs), which are insulating glass composed by glass plates entrapping a cavity filled with air sealed by perimeter metallic spacers usually made in aluminium or steel. The structural function of the glass plates is to withstand the external actions, whereas the role of the gas is correlated to the insulation of heat and noise in buildings, trains, ships, etc. [14]

Figure 1 shows a real photo of buildings made by insulating glasses [8]. This photo helps to illustrate not only the architectural beauty but also the utility and sustainability of the proper use of insulating glass in line with "building resilient infrastructure" in accordance with the 2030 Agenda for Sustainable Development [2].

**Fig. 1.** University of Toledo Engineering Building Toledo, Ohio, USA (retrieved from [8]).

The transmission of the external loadings between the glass plates, through the gas and spacers, is called "load sharing" phenomenon. In literature many studies show limited structural analysis in particular regarding this phenomenon, which in more advanced models has been better studied to estimate the out-of-plane deflections of the glass plates [4]. Other studies analyse this phenomenon also accounting for the buckling and the performance of the laminated DGUs [14, 15].

The best model that considers well this phenomenon is the Betti's analytical method (BAM), which has been validated experimentally in [3]. In [13], it is stated that the BAM is a "promising procedure that allows consideration of a wider range of geometries and loading configurations" and it is "an approach may be added in the final version of the Eurocode".

In this paper, the k-means algorithm has been used to find a direct method to “clean” the results of mechanical laboratory tests, which are usually raw outputs that they must be treated in analytical or visual way by filtering several data. In this sense, the paper is about post-processing laboratory measurements of deformed glass with greater accuracy than manual approaches. Here the out-of-plane deflections and elastic stiffnesses of the IGU glasses supported on four sides under linear and concentrated loads have been analysed. The results have been compared to that validated in [3].

The primary objectives of the k-means algorithm are: (i) to create clusters of cases or variables with high internal similarity and low similarity between groups; and (ii) to reduce data complexity in order to derive meaningful and actionable insights. In this sense, this algorithm can be very useful to treat in a more direct and quick way the outputs from experimental laboratory tests. To the best of the authors’ knowledge there are no studies where the k-means algorithm has been used for this purpose for DGUs.

## 2 Materials and Methods

### 2.1 Collected data

From previous published studies by the authors regarding the mechanical analyses of DGUs [3, 4] it was possible to collect several experimental data to be used for the present paper. In this sense, previous studies are used here as references to provide a new method for processing results from laboratory tests.

In [3, 4], the BAM model inserted in the new Eurocode [5] has been validated hence they can be considered as valid results. In particular, the collected data regard vertical deflections of DGUs under concentrated and linear loadings.

Figure 2 shows an example of the mechanical test of a DGU indicating the wooden box, which has been built to simulate the mechanical supports, the hydraulic dynamic actuator and load cell to produce and control the external load, and the metal bar to transfer the load in a linear way to the DGU. For a concentrated load a small wooden plate was used instead of the bar.

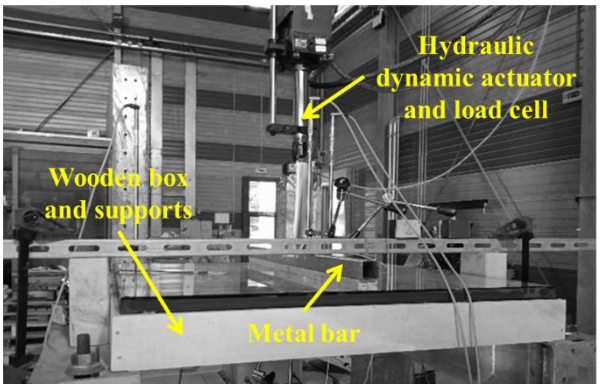



**Fig. 2.** Photo of the mechanical test of a DGU.

All studied DGUs are supported on 4 sides and are subjected to concentrated (about 1.0 kN) and linear (about 2.0 kN) loads. These values were sufficient to estimate the load-displacement curves hence the elastic stiffness avoiding brittle cracks (for more details see [3]).

For the purpose of the present study, the values of interest are considered as "raw values", which are treated here by the k-means algorithm to find possible correlation. These raw values, regarding the different DGUs (3 rectangular and 1 square), which are made by a superior and an inferior glass, aluminium spacers and air, are listed in Table 1.

**Table 1.** Used parameters for the analyses [3, 5].

| Parameter | Value |
|---|---|
| Glass thickness | 6.0 – 8.0 mm [a] |
| Gas cavity thickness (air) | 12.0 – 18.0 mm [a] |
| Side length | 0.50 – 2.0 m [a] |
| Young's modulus of the glass and aluminium | 70.0 GPa |
| Poisson ratio of the glass | 0.22 |
| Poisson ratio of the aluminium | 0.30 |
| Density of the glass | 2500.0 kg/m$^3$ |
| Density of the aluminium | 2700.0 kg/m$^3$ |
| Density of the gas | 1.18 kg/m$^3$ |

[a] The dimensions of the rectangular (0.50×2.0 m) DGUs are: 8×18×8 mm, 8×12×8 mm, 6×18×6 mm; the dimensions of the square (1.0×1.0 m) DGU are: 8×18×8 mm.

The values in Table 1 follow classical manufacturing, however some differences are possible in different world markets. Note that in practical applications for commercial façade units and residential windows, sizes for glasses around 1800.0 × 900.0 mm and

1250.0 × 750.0 mm, respectively, can be usually found in China [6, 7]. However, due to the great demand of this type of structures and more complex and futuristic solutions other sizes can be also manufactured, as for instance, 500.0 × 2000.0 mm and 1000.0 × 1000.0 mm in Europe. However, the method proposed in this study can be valid for any size.

## 2.2 Methodology

The main goal of the k-mean algorithm is to produce groups with a high degree of similarity and to reduce the complexity of the data treatment. For brevity, here only the main procedures are explained since in general this methodology is well known.

Firstly, it is necessary to collect relatively homogeneous data as shown in previous section. This initial step allows to provide more reliable outputs. The used values in the present paper refer to similar DGUs in term of mechanical characteristics hence they can be correctly used for this algorithm.

Then, the values are plotted in pairs, since the k-means algorithm adopts the Cartesian coordinates, providing groups of points (or clusters). By choosing the distance between the centre of a cluster, $c_i$, and general points, $p_{ij}$, with i = {1, 2, 3, …, k} and j = {1, 2, 3, …, n}, where k indicates the cluster numbers, the estimation of the Euclidean distance, E, and its minimum value can be obtained by, respectively:

$$E = \sum_i^k \sum_j^n \left\| p_{ij} - c_i \right\|^2 \quad (1)$$

$$\min\{E\} = \min\left\{ \sum_i^k \sum_j^n \left\| p_{ij} - c_i \right\|^2 \right\} \quad (2)$$

where the parameter n represents the number of cases included in a cluster k providing specific $p_{ij}$ points (note that two subscripts are necessary since, as mentioned, the Cartesian coordinates in a plane are used).

The position of $c_i$ is defined in a random way and then is adjusted. Mathematica software [9] provides automatic solutions under multiple iterations that are intrinsically computed. This software considers that the elements with high E distance are low similarity. Finally, for each $c_i$ value, the condition $dE/dc_i = 0$ must be verified.

In general, this algorithm is not difficult to be implemented also concerning the choosing of k-number. Here the value of k = 3.0 has been adopted since in previous studies (e.g., [10]) it has been noted that in general this number of clusters is sufficient to understand the phenomenon; this is because clusters 1 and 3 can serve as external limits thus cluster 2 can provide a reliable mean trend.

However, the analytical method to define the k-value consists in plotting dE values vs. k numbers and finding the “knee” of the obtained curve. The point where the curve changes its slope represents an optimum k value as explained in [1, 11]. In [1] the empirical effectiveness of this method has been demonstrated not only for choos-

ing the number of clusters but also for identifying underlying structure, on a wide range of simulated and real-world data sets.

## 3 Analyses and Results

By applying the described method, the raw values obtained by experimental laboratory tests can be plotted by using three clusters. To show the effects in applying the k-means algorithm, in Fig. 3a) the out-of-plane deflections (or vertical deflections) of the glasses in function of the applied concentrated loadings are plotted by using non clustered values (i.e., raw values), whereas Fig. 3b) shows the clustered values highlighting three groups. The deflection-load curves refer to the top central rectangular DGU under concentrated load (for brevity only one example is shown).

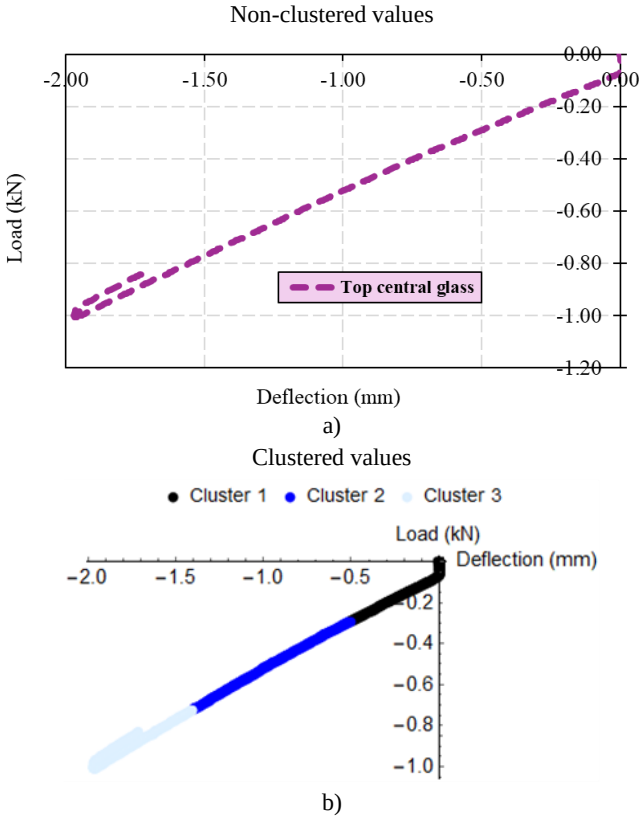



**Fig. 3.** Deflection values of the glass under concentrated load for a) non-clustered and b) clustered values.

In general, all non-clustered curves have some inconsistent parts, e.g., non-linear branches, the presence of discharge curves (like Fig. 3b)), curves that start at different times, etc.; hence these results could justify the use of the k-means algorithm to adjust and "clean" them. All tests provided a total of about $35.0 \times 10^3$ data, which must be treated to obtain useful outputs; in this sense, the use of the k-means algorithm can be beneficial for the purpose of the present study.

The manual adjustments have been made in terms of i) signs, where bottom and top deflections are considered positive if directed downward; ii) axis origins, where the deflections are null when the load is null; and time phases, where the deflections and load start at the same time; and iii) the unloading curves, from the maximum load to zero, have been deleted. Therefore, the results under these adjustments provide the validated values published in [3] and called in the present paper as "exact" values. Thus, they are compared with the clustered values obtained by Eqs. (1)-(2) in this analysis to appreciate the efficiency and goodness.

Figure 4 shows some comparisons in terms of probability density function (PDF) curves with respect the elastic stiffness (i.e., load/displacement ratio) for some DGUs under linear and concentrated loads. The k-means algorithm considers a cluster as a probability for each case, based on the mean value, $\mu_i$, standard deviation, $\pm\sigma_i$, and its PDF; for this reason, this type of representation is consistent in this way.

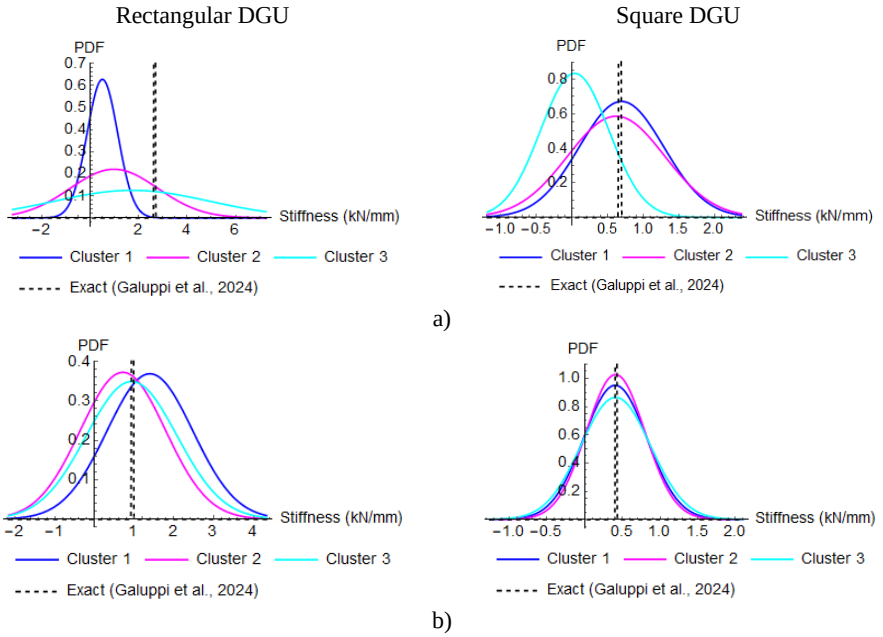



**Fig. 4.** PDF curves in function of the stiffness of the superior glass for a) linear and b) concentrated loads (some results).

Results for three clusters are shown and compared with the mentioned exact results; in this sense 48 stochastic analyses have been generated. The exact stiffness follows a linear-elastic behaviour, whereas the stiffness by cluster analyses could include some non-linear phenomena; however, the use of this algorithm also serves to find in a direct way the possible elastic stiffness at any point of a DGU.

The cluster that provides a $\mu_i$ value more similar to the exact value is assumed as the best cluster. Some clusters have a sharpened shape with a low $\pm\sigma_i$ value (i.e., low statistical error), however they could be not a better solution. In general, the deflections measured at the superior glass under concentrated load appear less scattered, with respect linear load, probably due to a direct and immediate transmission of the concentrated load up to the measuring point. In fact, the linear load can create strong stress concentrations in an extended part of the glass altering this transmission. Actually, in [3], the best agreement has been noted for superior deflections under concentrated load.

Some differences have been also registered between superior and inferior deflections of the glasses. In an DGU the transmission of the load between both glasses can be altered with different "velocity" due to influences of the mechanical properties of the internal gas and metal spacer. In a representation like Fig. 4 it is difficult to quantify separately the influence of the gas and/or spacers, as well as the geometric effects, since this algorithm follows a process purely geometric; however, both effects play an important role thus they must be considered. In general, the best results are provided by cluster 2 with a mean value of $1.01 \pm 0.06$.

Figure 5 shows all results of the exact solution (y-axis) with respect the approximated solutions (x-axis) obtained in this study regarding the stiffness of the DGUs. The correlation appears satisfactory with a R-squared value, $R^2$, of 0.97 – 0.99, showing the goodness of the algorithm, hence it appears a base for possible estimations for new and future DGUs. The loads have been divided due to their different influence on the DGUs, hence in Fig. 5a) the curves refer to the concentrated loads, whereas Fig. 5b) to the linear loads. This allows to obtain some trend lines and relations more useful to estimate a possible stiffness for DGUs.

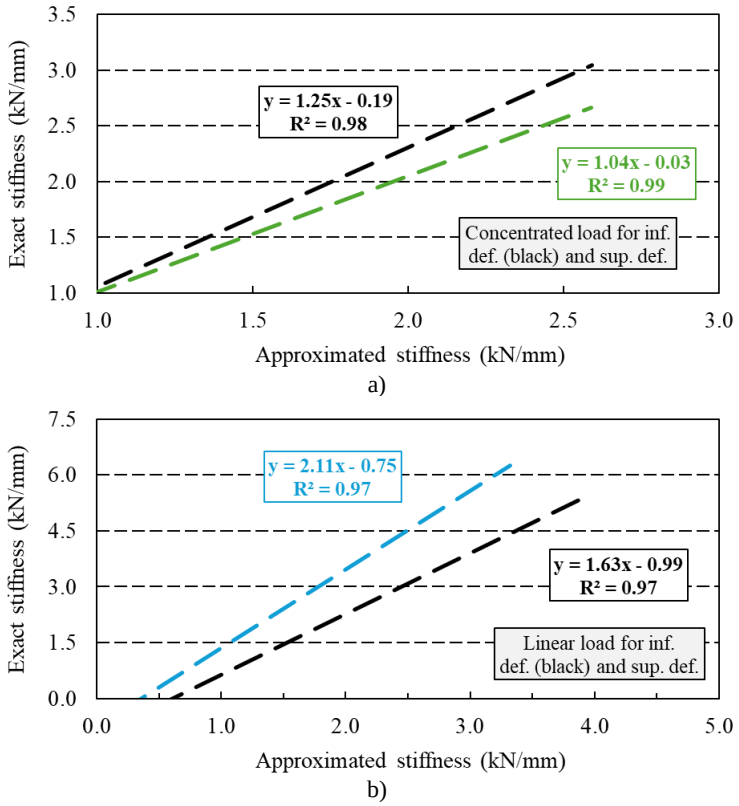



**Fig. 5.** Relation between exact and approximated stiffness for a) concentrated and b) linear load.

By using four relations in Fig. 5, it is possible to estimate the out-of-plane deflections of the glasses. These relations should represent the innovative and interesting aspect of this methodology. Note that the relations calculate the stiffness which, being linear elastic, is direct correlated to the deflection. The estimated relations are valid only for positive values (i.e., $y \geq 0$) to respect the physical meaning of the mechanical model.

Finally, Fig. 6 shows the mapping of the superior/inferior and central/later deflection points for some DGUs. Figure 6a) refers to a rectangular DGU, whereas Fig. 6b) to a square DGU. The exact values are indicated in black since they are calculated in previous studies and they are very similar to the clustered values obtained in this paper (mainly for cluster 2), whereas new values are shown in orange.

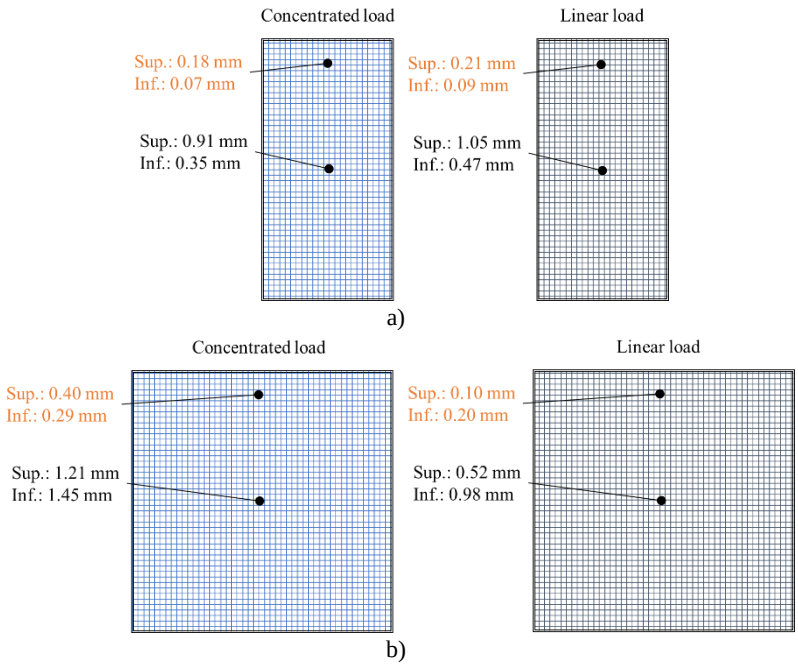



**Fig. 6.** Mapping of the deflections for a) rectangular and b) square DGUs (some examples).

The obtained values are very small; however, this approach is valid for any value under the linear-elastic model thus under the mechanical hypothesis of small deformations. In general, due to the brittle behaviour of the glass, small values are usually registered, and they represent the values of interest for design.

Results show that instead of using manual and analytical approaches to clean and adjust a-posteriori the mechanical laboratory tests, by using k-means algorithm and the estimated approximations it is possible to find good values for stiffness and out-of-plane deflections for a preliminary design.

## 4 Conclusions

In this paper the k-means algorithm applied for DGU glasses has been used to simplify the traditional mode to treat a-posteriori the results obtained from experimental laboratory tests. In particular here the out-of-plane vertical deflections thus elastic stiffnesses of square and rectangular DGUs were analysed. From mechanical tests, developed in previous studies, several data have been collected and treated by k-means algorithm. It appears that this approach is quick and quite reliable, however it is valid under specific hypotheses and conditions explained in this paper.

The k-means algorithm has been developed numerically by parametric software, obtaining about 50.0 statistical results. In general, cluster 2 appears the best cluster to be

considered. Linear trend and relations for concentrated and linear loadings have been obtained to estimate a possible elastic stiffness for DGUs. Thus, a mapping of out-of-plane deflections of the superior and inferior glass has been plotted a-posteriori.

Instead of using manual and analytical approaches to clean and adjust the mechanical laboratory tests, by using k-means algorithm and estimated approximations it is possible to find good values for stiffness and deflections, which are the key parameters for designing glasses.

## Acknowledgements

The authors acknowledge the Portuguese company SOSOARES and the ITECONS institute for prototyping the insulating glass units and carrying out the laboratory tests. The authors acknowledge the University of Salamanca (USAL), Spain, for providing a Wolfram Mathematica license and their institutions for covering the publication access fees, when applicable, required to download the referenced papers.

## References

1. Sugar, C.A., James, G.M., Finding the number of clusters in a dataset: an information-theoretic approach, J Am Stat Assoc 98, 750-763, 2003.
2. United Nations (UN), Transforming Our World: The 2030 Agenda for Sustainable Development, Technical Report, New York, USA, p. 35. A/RES/70/1. Accessed on 02/2026: https://sdgs.un.org/goals
3. Galuppi, L., Zacchei, E., Esteves, M., Ferrão, J., Simões, N., Experimental validation of the Betti's Analytical Method for Double Glass Units, Engineering Structures 315, 1-14, 2024.
4. Galuppi, L., Zacchei, E., Esteves, M., Simões, N., Double Glass Units simply supported on two sides: Analytical-numerical modelling and experimental testing, Engineering Structures 334, 1-14, 2025.
5. CEN/TG 250/SC 11, CEN/TS 19100 Design of glass structures, standard, 2021.
6. Wang, Z., Liu, J., Li, D., Yang, K., Chen, M., Wang, C., Experimental and numerical study on load-bearing performance in triple-glazed insulating glass units, Construction and Building Materials 418, 1-15, 2024.
7. Wang, Z., Liu, J., Li, D., Li, J., Wang, C., Yang, B., Ouyang, S., Study on the effect of geometric nonlinearity on uneven load distribution in multi-glazed insulating glass units, Journal of Building Engineering 111, 1-21, 2025.
8. Glass Metal Craft Company, Wixom, MI, USA. https://glassandmetalcraft.com/gmc_projects/university-of-toledo-engineering-building/
9. Wolfram Mathematica, version 12.0, Wolfram Research, Champaign, IL, USA, 2019.
10. Zacchei, E., Wang, Z., Brasil, R., K-means-based postprocessing of laboratory measurements of deformed insulating glass units, Journal of Architectural Engineering, 1-10, 2026.
11. Salvador, S., Chan, P., Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms, Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 1-9, 2004.

12. Afshoon, I., Miri, M., Mousavi, S.R., Combining Kriging meta models with U-function and K-Means clustering for prediction of fracture energy of concrete, Journal of Building Engineering, 35, 1-16, 2021.
13. Feldmann, M. et al., The new CEN/TS 19100: Design of glass structures, Glass Struct Eng 8, 317-337, 2023.
14. Louter, C., Belis, J., Veer, F., Lebet, J.P. 2012. Structural response of SG-laminated reinforced glass beams; experimental investigations on the effects of glass type, reinforcement percentage and beam size, Engineering Structures, 36: 292-301.
15. Bedon, C., Amadio, C., Buckling analysis and design proposal for 2-side supported double Insulated Glass Units (IGUs) in compression, Engineering Structures 168, 23-34, 2018.

# Simulation-based Comparative Study of PROFIBUS and PROFINET Topologies

Nurtakin Guliyeva Shukurova[1][0009-0004-9593-2595]

Azerbaijan Technical University, Baku, Azerbaijan

**Abstract.** The study investigates the impact of various network topologies used in the PROFIBUS and PROFINET communication protocols, which are widely used in industrial automation systems, on performance and system resilience. The study aims to determine the possibilities for optimizing the topology selection in Industrial Control Systems (ICS) networks in terms of connection reliability and continuity.

In the initial phase of the study, the bus, star, tree, and ring topologies were modeled in the GNS3 simulation environment and tested under scenarios such as cable breaks, network load, and changes in traffic intensity. The results showed that in the line topology, a 100% link loss was observed during a cable break, whereas in the ring topology, link continuity was maintained thanks to the alternative transmission path.

In the next phase, to address the identified weaknesses, the implementation of redundant structures, network segmentation, optimization of the switch architecture, and the creation of alternative transmission paths have been proposed at the theoretical and simulation levels. It should be noted that the optimization measures have not been fully implemented in a real industrial environment; they have only been evaluated within the framework of modeling and analysis.

Based on the simulation results, the application of redundant and distributed structures has the potential to increase link continuity during failures and can enhance the overall reliability of the system. At the same time, the use of segmentation and alternative transmission paths can reduce network latency and minimize packet loss. The obtained results provide a methodological contribution to the scientific justification of topology selection during the design of industrial networks and to ensuring the continuity of production processes. Furthermore, the practical implementation and experimental verification of the proposed optimization mechanisms on real hardware is considered appropriate for future research.

Although evaluated solely through GNS3 simulations, which may not fully capture physical-layer effects in real ICS environments, the proposed optimizations show promise; future research should validate these on actual PROFIBUS/PROFINET hardware with PLCs.

**Keywords:** Industrial Control Systems network, PROFIBUS, PROFINET, GNS3, topology optimization.

## 1 Introduction

The Industry 4.0 concept has accelerated the digital transformation of manufacturing systems over the last decade and has become one of the main directions of scientific research. Extensive literature reviews show that this concept is not limited to

merely increasing the level of automation, but also involves the integration of Cyber-Physical Systems, the Industrial Internet of Things, and communication infrastructure that ensures real-time data exchange [1]. In this regard, the performance, latency, and resilience of industrial networks are critical in an Industry 4.0 environment. Therefore, the architecture and topological structure of industrial networks directly impact the effectiveness of manufacturing systems.

Reliable and continuous data transmission in industrial automation systems is one of the key conditions for the stable and safe management of production processes. The PROFIBUS and PROFINET protocols, which are widely used in modern industrial networks, offer high-speed and deterministic communication capabilities [2], [3]. However, improper selection of the network topology can lead to connection loss, increased delays, and reduced system resilience [4]. In particular, in bus-type topologies, a single point of failure can cause the entire system to stop [5].

PROFIBUS is a traditional fieldbus that primarily operates over serial communication, while PROFINET is a modern, Ethernet-based industrial network technology. The correct and optimal application of these technologies is directly related to the structure of the chosen topology. Although performance analysis of industrial networks has been conducted in existing studies, the optimization of various topologies in a simulation environment and their evaluation under fault scenarios have been investigated to a limited extent.

In this paper, industrial network topologies are modeled in the GNS3 simulation environment, tested under failure and traffic load scenarios, and the impact of optimization mechanisms is analyzed.

**Research Questions:**

RQ1: Which topology minimizes packet loss under cable break scenarios?
RQ2: How does redundancy reduce latency in network topologies?

## 2. PROFIBUS and PROFINET

PROFIBUS (Process Field Bus) is a fieldbus technology that provides communication between PLC (Programmable Logic Controller) devices, sensors, drives, and other systems in industrial control systems. This technology, which complies with the IEC61158 standard, also has DP (Decentralized Peripherals) and PA (Process Automation) variants. The DP variant is used for systems such as machine automation. PA, on the other hand, creates a safe physical layer for use in areas with an explosion hazard. Although it has similarities to fiber optic and infrared cables, this technology, which primarily uses cables based on the RS-485 standard, ensures the operation of industrial control systems and provides compatibility between devices from different manufacturers [7].

PROFINET is an industrial network standard that provides connectivity between IT and OT. It is an Ethernet-based protocol used in industrial automation that operates in

real time, providing secure and higher-quality production. It facilitates data exchange between controllers, such as PLCs, and devices, and can use protocols like TCP/IP and UDP/IP[6].

PROFIBUS primarily operates over serial communication, while PROFINET is Ethernet-based. The performance of PROFIBUS and PROFINET networks can be compared using several key parameters. PROFIBUS DP is a serial network protocol with a maximum speed of 12 Mbit/s and supports a maximum of 32 nodes per segment, expandable to a total of up to 126 nodes. PROFINET, on the other hand, is an Ethernet-based protocol with speeds of 100 Mbps and higher, which ensures deterministic behavior with real-time (RT) and isochronous real-time (IRT) mechanisms. For this reason, PROFINET is suitable for larger and more flexible topologies, while PROFIBUS is practical for smaller and more stable systems[6], [7].

**Table 1.** Comparative Technical Analysis of PROFIBUS and PROFINET Communication Protocols.

| Parameter | PROFIBUS DP | PROFIBUS PA | PROFINET RT | PROFINET IRT |
|---|---|---|---|---|
| Physical Layer | RS-485 | MBP (IEC61158-2) | Industrial Ethernet | Industrial Ethernet |
| Max Data Rate | 12 Mbit/s [3] | 31.25 kbit/s [2] | 100 Mbit/s [8] | 100Mbit/s (isochronous) [8] |
| Topology | Line, Tree | Line | Star, Line, Ring | Star, Ring |
| Real-Time Capability | Deterministic | Deterministic | Soft Real-time [2] | Hard Real-time [2] |
| Max Nodes | 126 devices[3] | 32 per segment [2] | >100 devices [8] | >100 devices [8] |
| Redundancy | Limited | Limited | MRP Support [8] | Advanced Redundancy [8] |
| Security Features | Basic | Basic | Advanced Security Concepts [9] | Advanced Security Concepts [9] |

Table 1 provides a comparative overview of PROFIBUS and PROFINET technologies based on official specifications and the latest scientific literature. PROFIBUS DP supports data transmission speeds of up to 12 Mbps, while PROFIBUS PA operates at a speed of 31.25 kbps using the MBP physical layer [3]. In contrast, PROFINET has an Ethernet-based architecture and provides speeds of up to 100 Mbit/s, as well as real-time (RT) and isochronous real-time (IRT) capabilities [8]. Furthermore, PROFINET offers advanced redundancy mechanisms, such as the Media Redundancy Protocol (MRP), and enhanced security concepts for industrial communication environments [9].

## 3. TOPOLOGIES FOR PROFINET AND PROFIBUS

### 3.1 PROFIBUS DP Topologies

PROFIBUS DP is a serial fieldbus technology that provides deterministic data transmission between distributed field devices and a central control system in industrial automation systems. According to the official system description, a line topology is primarily used at the RS-485 physical layer [2].

In a line topology, all devices are connected in series on the same cable, and it is mandatory to place terminators at both ends of the segment. Incorrect selection or placement of terminators can cause signal reflection and communication errors.

Real-time analyses conducted by Eduardo Tovar and Francisco Vasques have shown that although PROFIBUS DP provides deterministic time behaviour in a line topology, a single point of failure (cable break or device failure) can affect the operation of the entire segment. For this reason, tree-type extensions via a repeater are implemented in large systems. Thus, although the linear topology is considered a simple and economical solution for PROFIBUS DP, it has limitations in terms of fault tolerance.

### 3.2 PROFIBUS PA Topologies

PROFIBUS PA is primarily intended for the process industry (chemical, oil and gas, and energy sectors) and uses the IEC 61158-2 MBP physical layer [2].

In PROFIBUS PA networks, line, star, tree, and, in some cases, ring topologies can be used. The tree topology is considered more practical in process industries due to the connection of widely distributed sensors and actuators.

J. P. M. D. P. Paiva et al. have shown that while PROFIBUS PA is more suitable for stable, low-speed applications compared to PROFINET PA, it has limitations in terms of scalability and flexibility.

Although the star topology allows for the isolation of local failures, the failure of the central point can cause a complete loss of communication.

### 3.3 PROFINET Topologies

PROFINET is an Ethernet-based industrial network technology and supports high speeds (100 Mbit/s and higher) and a combination of various topologies [2].

PROFINET supports star, line, tree, ring, and hybrid structures. In a star topology, all devices are connected through a central switch. This structure allows for the isolation of local failures, but a failure of the central switch causes a network outage.

The line topology has a simpler structure and can be built without a switch in some cases. However, a failure of one device can cause subsequent devices to lose connectivity. The ring topology, on the other hand, provides high fault tolerance through the Media Redundancy Protocol (MRP). Alexander Walz and his co-authors have shown that in PROFINET networks, redundancy mechanisms increase not only performance but also safety. Furthermore, the PROFINET RT and IRT modes directly influence the choice of topology. The IRT mode is optimized for applications requiring high-precision synchronization.

### 3.4 Comparative Analysis and Optimization Aspect

The choice of topology is not only a matter of physical configuration but also of optimizing network performance.

In PROFIBUS networks, a linear topology offers the advantage of low cable consumption and a simple structure, but it has a high risk of a single point of failure. PROFINET, on the other hand, allows for more flexible and redundant structures due to its Ethernet-based architecture.

Thus, as noted by Eduardo Tovar and Fernando Vasques, in serial fieldbus architectures, a deterministic transmission mechanism ensures predictable latency, but fault tolerance and structural redundancy remain limited in this type of network.

Andreas Walz and his co-authors, in turn, show that in Ethernet-based industrial networks, redundancy protocols, alternative routing mechanisms, and advanced security concepts significantly enhance system resilience and cybersecurity.

In addition, J. P. M. D. P. Paiva and colleagues emphasize that in the modern process industry, the choice of network topology should be based on the application's scale, the level of real-time requirements, and reliability metrics.

In this regard, the following technical parameters must be systematically evaluated during optimization: latency, packet loss, redundancy mechanism, recovery time in case of failure, as well as cable length and segment limits.

Thus, while standardized PROFIBUS networks with a linear architecture provide a cost-effective and technically efficient solution for small and medium-sized automation systems, in large-scale industrial environments requiring high reliability, PROFINET-based ring or hybrid topologies may be a more optimal choice. This choice is considered a strategic approach aimed at ensuring the network's resilience, continuity, and future scalability.

## 4. METHODOLOGY

The primary prerequisite for network optimization is the selection of an appropriate topology. For example, when selecting the topology for a Profibus network, the extent of the system's coverage area, the number of devices to be connected, cable distances, electromagnetic interference, and the system's reliability requirements must be taken into account. For example, while a bus topology is sufficient for small systems, tree or star topologies are more suitable for larger systems.

The optimization of the topology was evaluated within a simulation environment, not in a real industrial setting, through modeling. The selection criteria were determined based on theoretical literature and official technical specifications, but the practical implementation was only carried out on a virtual network model.

The GNS3 platform was used as the simulation environment. Although this platform allows for the emulation of the functional behavior of real network devices, it does not fully reflect physical environment factors (e.g., electromagnetic interference,

the effect of actual cable length on signal attenuation, or industrial-grade hardware delays).

In this environment, routers, switches, and other network devices were configured, and traffic analysis and packet-level observations were conducted. The main reason for selecting this platform is that it allows for the analysis of the dynamic behavior of industrial Ethernet architectures, particularly PROFINET-based topologies.

In the simulation phase, line, star, tree, and ring topologies were implemented for PROFINET. The Media Redundancy Protocol (MRP) mechanism was implemented in the ring topology. For PROFIBUS, a bus topology was modeled. An objective comparison was ensured by applying the same number of virtual devices and a similar traffic load for each topology.

The following test scenarios were executed as part of the study: physical cable cut (link failure), adding additional traffic load to the network, changing the transmission rate, and a comparison of the redundancy mechanism in active and disabled states. During the cable cut scenario, the network's recovery time was measured, while in the additional traffic scenario, latency and packet loss metrics were analyzed.

The following metrics were used for performance evaluation: end-to-end delay (ms), packet loss percentage, network recovery time (ms), traffic load, and connection duration. The total delay was evaluated as the sum of the following components: transmission delay, propagation delay, and processing delay. This approach allowed for an objective comparison between the topologies.

A comparative analysis of the simulation results showed that although the linear topology is superior in terms of minimal cable usage and simple structure, the failure of a single node can disrupt the operation of the entire segment. Star topology allows for the isolation of local failures, but the failure of the central switch leads to a complete loss of connectivity. The tree topology offers a balanced solution for larger areas. The ring topology, on the other hand, provides high fault tolerance through the MRP mechanism, allowing for the use of an alternative route during a failure and minimizing recovery time.

Thus, the simulation analysis conducted has shown that for small and medium-sized systems, simple linear structures may be sufficient, but in industrial environments where high reliability and durability are required, redundant ring or hybrid topologies are considered more optimal. The choice of topology is not just a matter of physical layout but also a strategic decision in terms of optimizing performance and reliability.

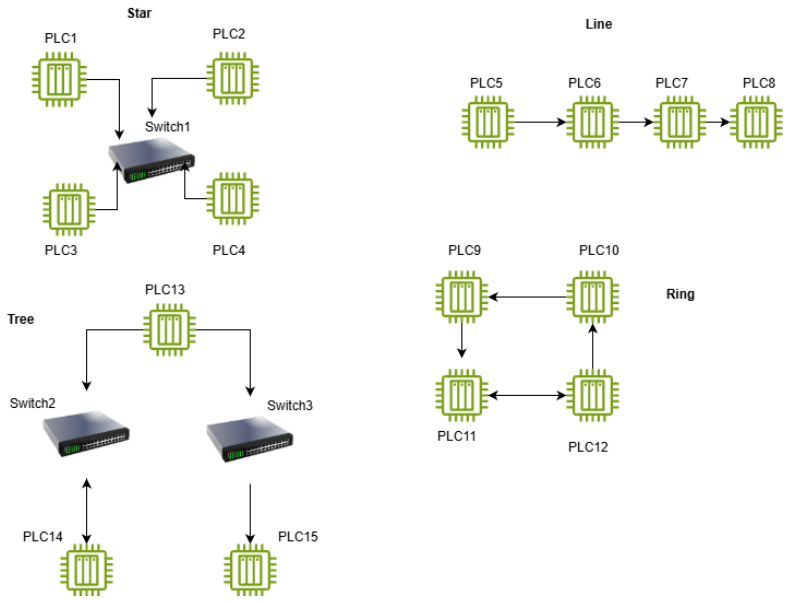



**Figure 1. Star, ring, line, bus topologies**

Various scenarios were tested during the simulation process: the physical cutting of the cable, adding an additional load to the network, and changing the transmission speed. As a result of these tests, metrics such as latency, packet loss, and link stability were observed.

Tests conducted on a star topology set up in the GNS3 simulation environment have shown that the network operates stably and with low latency under normal conditions. The average latency values for ICMP requests sent from PC1 to PC3 were in

the 1–2 ms range.This result aligns with the performance characteristics of Layer 2-based Ethernet networks. However, during the testing process, in some cases, the latency values rose above 200 ms, and short-term packet loss was observed. This situation is not due to the network architecture but rather to resource shortages in the simulation environment (CPU load, virtual interface latency, etc.). Such high latency values are not typical for properly configured Ethernet-based PROFINET networks in a real industrial environment.

The main advantage of a star topology is that the failure of one of the end devices does not affect the operation of the other devices. However, a failure of the central switch causes the entire network to fail. This characteristic is considered critical for reliability in industrial networks and necessitates the implementation of additional redundant mechanisms.

High availability and deterministic transmission are crucial in industrial Ethernet-based systems, especially in PROFINET architectures. Research shows that while centralized topologies are effective in terms of performance, they pose a risk for critical processes without redundant structures[8].

### 4.1 Recommendations for Optimization

To eliminate the weaknesses of the star topology and increase industrial-level reliability, the following optimization methods are recommended:

1. Adding a Redundant Switch

Adding a second switch as an alternative to the central switch increases network availability. This approach helps eliminate the single point of failure issue. In industrial networks, this method is one of the key elements of a high-availability strategy.

2. Transition to a Tree Structure

Load balancing can be achieved by creating a tree structure by combining several star topologies. This approach provides more effective traffic distribution in large-scale industrial facilities [11].

3. QoS and Traffic Prioritization

In a real industrial environment, it is possible to assign a priority for critical control traffic on PROFINET networks. This mechanism minimizes latency and ensures real-time transmission. In particular, the RT and IRT mechanisms play a crucial role in maintaining deterministic performance [10].

4. Ring and MRP Mechanism

Although the Media Redundancy Protocol (MRP) is primarily used in a ring topology, a similar level of reliability can be achieved with a redundant switch architecture. MRP ensures network resilience by activating an alternative route within milliseconds during a link failure.

## 5. RESULTS AND DISCUSSION

The simulation conducted in the GNS3 environment provided a comparative quantitative analysis of PROFIBUS and PROFINET topologies under identical conditions, including node count and traffic intensity. The performance was evaluated based on latency, packet loss, and recovery time across two primary states: normal operation and link failure scenarios.

**Table 2. Performance Comparison Under Normal Conditions**

| Topology | Latency | Packet Loss | Throughput | Stability |
|---|---|---|---|---|
| Star | 1-2 ms | 0% | High | Very high |
| Bus | 4-7 ms | 5-10% | Medium | Medium |
| Ring | 1-2 ms | 0 -2 % (low load) | Low-Medium | Low |
| Tree | 2-4ms | 0-5% | High | High |

### 5.1. Performance Under Normal Conditions

As summarized in Table 2, the Star and Ring topologies demonstrated the highest efficiency for PROFINET-based architectures. The Star topology yielded the lowest latency (1–2 ms) and 0% packet loss, confirming its high transmission stability for real-time industrial data. Conversely, the Bus (Line) topology exhibited higher latency (4–7 ms) and a packet loss rate of 5–10%, primarily due to increased propagation delays and the lack of a centralized switching mechanism.

**Table 3.Performance Under Fault and Stress Scenarios**

| Topology | Scenario | Packet Loss | Recovery Time | Result |
|---|---|---|---|---|
| Star | Link Failure | 20-40% (affected node only) | 1–3s (Fast recovery) | Resilient |

| | | | | |
|---|---|---|---|---|
| Bus | Traffic Increase | 40-80% | N/A (Broadcast collisions) | Unstable |
| Ring | Congestion | 15-30 % | N/A (Queue buildup) | Degraded |
| Tree | Link Failure | 20-60% | N/A (Subtree isolation) | Partially Resilient |

### 5.2. Resilience and Fault Tolerance Analysis

The study's most significant findings occurred during the link failure (cable cut) simulations, detailed in **Table 3**.

- Line Topology: A single cable break resulted in 100% link loss for all downstream nodes, confirming its vulnerability as a single point of failure in industrial environments.
- Ring Topology: Communication remained uninterrupted during a link failure. The implementation of the Media Redundancy Protocol (MRP) allowed the network to activate an alternative transmission path within milliseconds, ensuring near-constant link continuity.
- Star Topology: This structure allowed for the isolation of failures, where packet loss (20–40%) was restricted only to the affected node, while the rest of the network remained operational with a fast recovery time of 1–3 seconds.

### 5.3. PROFIBUS-Specific Observations

In the PROFIBUS DP and PA simulations, stable communication was found to be highly dependent on physical layer parameters. Specifically, the correct placement of terminators at both ends of a segment and strict adherence to standardized cable lengths were identified as critical factors in preventing signal reflection and ensuring deterministic behavior. While the line topology is cost-effective for small-scale PROFIBUS systems, it lacks the inherent fault tolerance observed in PROFINET-based redundant structures

## 6. CONCLUSION

This study shows that in industrial networks, not only the selection of the protocol but also the establishment of the correct topology is of great importance. While a properly configured bus topology with suitable parameters for PROFIBUS ensures stable communication, star and ring topologies stand out as a more flexible and secure solution for PROFINET.

Optimized topologies not only increase the reliability of industrial processes but also play a crucial role in preventing system failures.

### 6.1 Limitations

This study relies on GNS3 for modeling, which excels in network-layer simulation but lacks an accurate representation of PROFIBUS's RS-485 physical layer or PROFINET's real-time constraints on industrial hardware. Tests used synthetic traffic (e.g., ICMP/ping), not native cyclic I/O data from PLCs. No real-world factors like EMI or temperature were modeled.

## 7. FUTURE WORK

This study primarily examined the performance and reliability of different network topologies in a simulation environment. In future work, we plan to validate and extend these findings on real industrial hardware. This includes implementing the proposed architectures using Siemens S7 PLCs, capturing and analyzing actual PROFINET traffic, conducting evaluations based on IEC 61784 standards, and performing hardware-level fault injection tests to assess real-world behaviour under realistic failure conditions.

Additionally, exploring artificial intelligence and machine learning techniques offers a promising direction. ML models could dynamically monitor traffic patterns, latency, and packet loss to proactively detect anomalies and enable automatic rerouting, thereby improving resource management, real-time decision-making, and overall system resilience in large-scale industrial networks.

## References

1. L. S. Dalenogare, G. B. Benitez, N. F. Ayala, and A. G. Frank, “The expected contribution of Industry 4.0 technologies for industrial performance,” Technological Forecasting and Social Change, vol. 135, pp. 383–394, 2018.

2.PROFIBUS & PROFINET International, "PROFIBUS System Description – Technology and Application," PI White Paper, Karlsruhe, Germany, 2020..

3. PROFIBUS & PROFINET International, "PROFINET System Description – Technology and Application," PI White Paper, Karlsruhe, Germany, 2021.

4. J. Li, X. Li, and J. Gao, "Exploring Self-Organization and Self-Adaption for Smart Manufacturing Complex Networks," IEEE Access, vol. 8, pp. 132001-132014, 2020. doi: 10.1109/ACCESS.2020.3009575

5. X. Jiang, H. S. Lohne, and M. Gidlund, "Time-Sensitive Networking Technologies for Industrial Automation in Wireless Communication Systems," IEEE Industrial Electronics Magazine, vol. 17, no. 1, pp. 36-47, March 2023. doi: 10.1109/MIE.2022.3162381

6.S. P. Kumar and S. S. Shajin, "A Comparative Study to Evaluate the Performance of Communication Protocols for Process Industry," in Proc. 2023 7th Int. Conf. Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1105–1110.

7. PROFIBUS Nutzerorganisation e.V., PROFIBUS Design Guideline, Order No. 8012_V129, 2020.

8.A. Walz et al., "Profinet Security: Selected Concepts for Secure Communication in the Automation Domain," in Proc. IEEE Int. Conf. Industrial Informatics (INDIN), 2023..

9. J. P. M. D. P. Paiva et al., "Use Case Comparison Between Profibus PA and Profinet PA Networks," in Proc. IEEE Latin American Symp. Circuits and Systems, 2025.

10. F. Depuydt and P. Saey, "Redundancy for PROFINET Networks," White Paper, KU Leuven, Dept. Electrical Engineering (ESAT), Ghent, Belgium, 2010.

11.M. Raţă, A. Gontean, and D. Moga, "Performance Evaluation of PROFIBUS Networks in Industrial Automation Systems," in Proc. IEEE Int. Conf. Industrial Informatics (INDIN), 2022, pp. 1–6.