Volume 1, Issue 2

2026

COMPUTATIONAL AND APPLIED SCIENCE

International Peer-Reviewed Open Access Journal

Independent International Scientific Journal

Tbilisi

2026

Computational and Applied Science (CAS Journal) is an independent, international, peer-reviewed open access scientific journal dedicated to the development and dissemination of innovative theoretical, methodological, and applied research across a broad range of computational and applied disciplines. The journal particularly emphasizes research in computational and applied mathematics, scientific and numerical computing, algorithms and data structures, artificial intelligence and machine learning, data science and analytics, computational modeling and simulation, computer science and information technologies, digital transformation, business and decision analytics, educational technologies, and other related interdisciplinary fields.The journal provides a platform for the dissemination of theoretical, methodological, and applied research that contributes to the advancement of computational and applied sciences and supports scholarly dialogue at the international level.

All manuscripts submitted to the journal undergo a rigorous peer-review process to ensure academic integrity, originality, and scientific relevance.

The authors are solely responsible for the content, accuracy, and originality of their articles. Opinions expressed in published papers do not necessarily reflect the views of the editorial board or the publisher.

Journal Title: Computational and Applied Science
Abbreviation: CAS Journal
Publication Type: Open Access
Peer Review: Double-blind peer review
Language: English
Publisher: CAS Journal
Country of Publication: Georgia

ISSN 3088-4152 (Online)

Editorial Team

Contents

AI Framework for Academic Progress Monitoring and Personalized Assignment Recommendation in Higher Education - Besiki Tabatadze, Sophio Khundadze, Ilia Botsvadze 5

Real-Time Presentation of Courier Movement in Web and Mobile Applications Using Socket.IO and Message Broker Architectures - Anri Morchiladze 19

AI at Half the Price: Harnessing Off-Grid Hydropower and Direct Cooling for Scalable Intelligence - Giorgi Mushkudiani 31

Mathematical Modeling of the Rheological Behavior of Asphalt Concrete for Smart City Road Infrastructure - Maka Mantskava, Giorgi Kuchava, Giorgi Chanturia, Shawn Moodley 49

Review and Analysis of Cyberattack Detection Methods in Corporate Networks - Ioseb Kartvelishvili, Giorgi Kuchava, Demetre Chakvetadze 58

Design and Evaluation of a Hybrid Georgian Voice Assistant for Public Services - Tamar Tutisani, Nino Topuria . 75

AI-Assisted Editorial Decision Making in Open Journal Systems (OJS): A Framework for Manuscript Scope Analysis and Reviewer Recommendation -Besiki Tabatadze, Teimuraz Sturua, Nika Tabatadze, Ekaterine Natsvlishvili 83

An AI-Driven Multimethod Framework for Inclusive Education: A Cross-Regional Survey Analysis - Mariam Chkhaidze, Levan Mateshvili 103

Urban Transport Systems Optimization Through Real-Time Data Analytics-Nikoloz Patatishvili, Besiki Tabatadze 121

Self-Inflicted Denial of Service (Self-DDoS) in Distributed Systems, Mechanisms and Mitigating Architectural Strategies: When a System "Attacks" Itself, An Analysis of Retry Storms, Thundering Herds, and Cascading Failure- Tengo Gelishvili, Giorgi Kuchava, Teimuraz Sturua 141

Morphological Classification of Tobacco Ash Using a Deep Convolutional Neural Network: From Sherlock Holmes's "monograph" to modern computer vision - Irine Imerlishvili, Giorgi Kuchava 149

Development of a Monitoring Dataset for Logic-Based Electromagnetic Field Risk Assessment - Lia Kurtanidze. 157

Artificial Intelligence and its Role in Crisis Management: Opportunities and Challenges of Technological Intervention -Ana Mazmishvili 172

Conceptual Model of a Hybrid Automated AI-Based Code Review System - Tea Todua, Giorgi Tsitlidze 189

AI Framework for Academic Progress Monitoring and Personalized Assignment Recommendation in Higher Education

Besiki Tabatadze

PhD (Applied Mathematics), Professor, European University, Georgian American University, Tbilisi, Georgia.

Email: *tabatadze.besik@eu.edu.ge.*

Sophio Khundadze

PhD, European University, Tbilisi, Georgia.

Email: *skhundadze@eu.edu.ge.*

Ilia Botsvadze

PhD, Associate Professor, European University, Tbilisi, Georgia.

Email: *ilia.botsvadze@eu.edu.ge.*

Abstract

The rapid development of Artificial Intelligence (AI) has significantly expanded its applications in higher education, particularly in the fields of learning analytics and personalized learning. Continuous monitoring of students' academic progress and providing personalized learning recommendations remain major challenges in higher education. This paper proposes an AI-assisted framework for academic progress monitoring and personalized assignment recommendation based on a Decision Tree classification model for predicting students' expected academic performance. The proposed approach integrates students' learning activities, formal assessment results, and historical instructor decisions to predict the expected grade category and automatically generate personalized assignments with an appropriate level of difficulty. A prototype of the framework was implemented in Python using the scikit-learn library and evaluated on a dataset of 50 students. Experimental results demonstrate that the proposed framework provides transparent and interpretable predictions, reduces the effort required for continuous monitoring of student progress, and supports the effective implementation of personalized learning. Rather than replacing instructor judgment, the proposed approach serves

as an intelligent decision-support tool that can be integrated into existing learning management systems and adapted to a wide range of higher education courses.

Keywords: Artificial Intelligence, Decision Tree, Educational Data Mining, Learning Analytics, Academic Progress Monitoring, Personalized Learning, Higher Education.

Introduction

Artificial intelligence (AI) is playing an increasingly important role in higher education by supporting data-driven educational decision-making and personalized learning. Advances in educational data mining and learning analytics have enabled institutions to collect, process, and analyze student performance data more effectively, providing valuable insights into learning processes and academic outcomes (Baker, 2010; Romero & Ventura, 2010; Siemens & Long, 2011). This study extends previous research on AI-assisted educational technologies and academic analytics (Tabatadze, 2024; Tabatadze & Khundadze, 2026), this study focuses on academic progress monitoring and personalized assignment recommendation.

Despite these developments, continuous monitoring of student progress remains a challenging task for university instructors. Student performance is typically assessed through multiple learning activities and formal assessments conducted throughout a course, requiring instructors to continuously evaluate diverse information in order to identify students who need additional academic support. In many cases, learning difficulties become evident only after formal assessments have taken place, reducing opportunities for timely intervention. Previous studies have therefore emphasized the importance of early prediction of student performance using educational data (Cortez & Silva, 2008; Kabakchieva, 2013; Marbouti et al., 2016).

Another challenge is determining the most appropriate learning activities once students requiring additional support have been identified. This often involves manually reviewing

previous course topics, assignment difficulty levels, and prerequisite relationships, making personalized recommendations both time-consuming and potentially inconsistent.

To address these challenges, this study proposes an AI-assisted framework for academic progress monitoring and personalized assignment recommendation based on a Decision Tree classification model. The framework integrates students' learning activities, formal assessment results, and historical instructor decisions to predict expected grade categories and automatically generate personalized assignments with appropriate difficulty levels. Unlike many existing approaches that focus solely on predicting academic performance, the proposed framework combines prediction and recommendation within a unified architecture that supports instructors throughout the learning process.

The proposed approach makes three main contributions. First, it introduces a checkpoint-based framework that continuously updates student performance predictions as new assessment data become available. Second, it integrates predictive analytics with a transparent recommendation mechanism that automatically generates personalized assignments according to students' predicted performance. Third, it presents an experimental evaluation of the proposed framework using real educational data and demonstrates its practical applicability through quantitative analysis and representative case studies.

2. Literature Review

Artificial intelligence (AI) has become an important component of higher education, supporting academic administration, teaching, automated assessment, educational analytics, intelligent tutoring systems, and early identification of at-risk students (Tabatadze, 2024; Tabatadze, 2026; Tabatadze & Khundadze, 2026). Across these applications, AI primarily serves as a decision-support technology that assists instructors while preserving human oversight over educational decisions.

Educational Data Mining (EDM) and Learning Analytics (LA) provide the methodological foundation for AI-based educational systems. EDM applies machine learning and data mining techniques to identify patterns in educational data and predict student performance (Baker, 2010; Romero & Ventura, 2010), while Learning Analytics focuses on collecting and analyzing learner-generated data to improve teaching and learning processes (Siemens & Long, 2011). Previous studies have demonstrated that student engagement and participation are reliable indicators of academic success and support timely educational interventions (Romero et al., 2013; Baker & Inventado, 2014).

Student performance prediction is one of the most widely studied applications of machine learning in education. Numerous studies have successfully applied supervised learning methods to identify at-risk students using historical academic and behavioural data (Cortez & Silva, 2008; Kabakchieva, 2013; Marbouti et al., 2016). Among these methods, Decision Tree classifiers are particularly attractive because they combine satisfactory predictive performance with high interpretability, enabling instructors to understand the reasoning behind each prediction (Breiman et al., 1984; Quinlan, 1986; Al-Barrak & Al-Razgan, 2016; Kotsiantis, 2012).

Another important research direction is personalized learning, where instructional content and learning activities are adapted to individual students' knowledge and learning needs (Brusilovsky, 2001; Chrysafiadi & Virvou, 2013). Previous studies have demonstrated the effectiveness of personalized diagnosis, adaptive recommendation systems, and individualized learning pathways in improving learning outcomes (Chen, 2011; VanLehn, 2011). In a related direction, (Tabatadze and Sokhadze, 2023) showed that statistical analysis of collected assessment points can inform targeted improvements to course content.

Although significant progress has been made, most existing studies address student performance prediction and personalized learning separately. Few studies integrate these functions into a unified framework that continuously monitors academic progress while simultaneously generating personalized assignment recommendations. The framework proposed

in this study addresses this gap by combining predictive analytics and adaptive assignment recommendation within a single checkpoint-based architecture that supports instructors throughout the learning process.

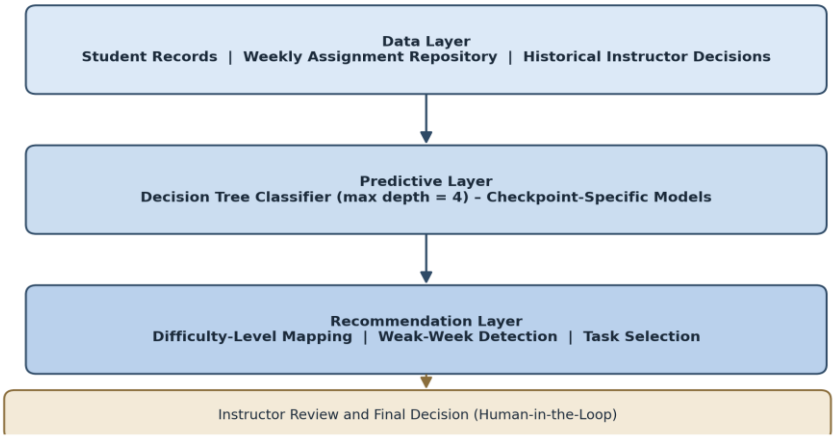
3. Proposed AI-Assisted Framework

3.1. Overall Framework

The proposed framework supports instructors in monitoring academic progress and generating personalized assignment recommendations at predefined checkpoints of a 17-week course, building on adaptive and student-modeling approaches that tailor instructional content to individual learners' estimated knowledge levels (Brusilovsky, 2001; Chrysafiadi & Virvou, 2013). The framework analyzes informal and formal assessment data to produce two outputs for each student to produce two outputs for each student: a predicted grade category and personalized assignment recommendations organized by topic and difficulty.

As illustrated in Figure 1, the framework comprises three components. The data layer stores weekly student records, a structured assignment repository, and historical instructor decisions used for model training. The predictive layer employs a Decision Tree classifier to estimate the expected grade category based on cumulative performance data available at each checkpoint. The recommendation layer translates the predicted grade into appropriate assignment difficulty levels, identifies topics requiring additional practice, and selects the corresponding assignments. The instructor retains full control over the final recommendations. The framework is implemented as a Python-based prototype using standard machine learning and data-processing libraries (Pedregosa et al., 2011).

Figure 1. Overall Framework Architecture

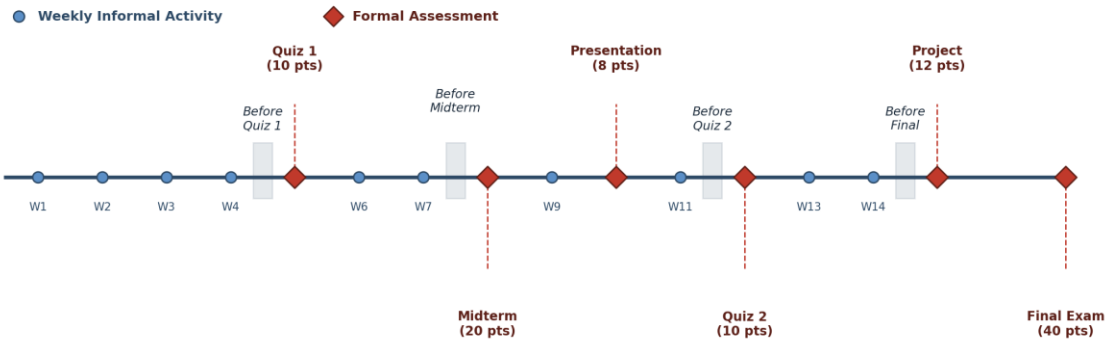


3.2. Dataset and Data Processing

The proposed framework operates on three categories of structured data: student records, the weekly assignment repository, and historical instructor decisions used to train the predictive model.

The student dataset used to evaluate the framework consists of 50 student records collected across a 17-week course. For each student, the dataset records weekly informal activity scores (on a 0-5 scale) for Weeks 1, 2, 3, 4, 6, 7, 9, 11, 13, and 14, together with the results of all formal assessment instruments: Quiz 1 (Week 5, maximum 10 points), the Midterm examination (Week 8, maximum 20 points), a Presentation (Week 10, maximum 8 points), Quiz 2 (Week 12, maximum 10 points), a Project (Week 15, maximum 12 points), and the Final examination (Week 17, maximum 40 points), for a total of 100 points. Figure 2 illustrates the course timeline and the points at which informal activity and formal assessment data are collected relative to each checkpoint.

Figure 2. Course Timeline and Data Collection Process



The weekly assignment repository organizes practical tasks for ten teaching weeks (Weeks 1, 2, 3, 4, 6, 7, 9, 11, 13, and 14), corresponding to individual course topics. Each week's assignments are categorized into four difficulty levels, Easy, Medium, Hard, and Complex, identified by task number. In addition, each week is annotated with its thematic prerequisite weeks (topic dependencies), reflecting the sequential structure of the syllabus.

A historical decision dataset, produced by the instructor for the Before-Midterm checkpoint, records the informal progress percentage, official assessment percentage, resulting grade category, and recommended tasks for each of the 50 students. This dataset serves as the training data (ground truth) for the predictive model at that checkpoint; analogous historical decisions are required to train the model at the remaining three checkpoints, and their collection is identified as a direction for future work in Section 6.

Grade categories follow a seven-level scale: A - Excellent (91-100%), B - Very Good (81-90%), C - Good (71-80%), D - Satisfactory (61-70%), E - Sufficient (51-60%), Fx - Fail, Retake (41-50%), and F - Fail (0-40%). Table 1 summarizes the experimental dataset used to evaluate the Before-Midterm checkpoint of the proposed framework.

Table 1. Summary of the Experimental Dataset

Dataset	Number of Records	Description
Student Records	50	Weekly informal activity scores and formal assessment results (Quiz 1, Midterm, Presentation, Quiz 2, Project, Final)
Weekly Assignments	10 weeks	Practical tasks organized by topic, difficulty level, and topic dependency
Historical Decisions (Before Midterm)	50	Instructor-labeled grade categories (7 levels, including Fx) and recommended tasks used for model training

3.3. Predictive Modeling of Academic Performance

The core predictive component of the framework is a Decision Tree classifier (maximum depth of four), trained separately at each of four course checkpoints: before Quiz 1, before the Midterm, before Quiz 2, and before the Final examination, following prior evidence that decision tree models can effectively predict a student's final grade from earlier-available performance indicators (Al-Barrak & Al-Razgan, 2016). Each checkpoint is associated with a cumulative feature set that grows as the semester progresses, reflecting the assessment information available to the instructor at that point in the course. Table 2 summarizes the feature configuration of each checkpoint.

Table 2. Feature Configuration at Each Course Checkpoint

Checkpoint	Cumulative Features Used for Prediction	Weeks Covered
Before Quiz 1	Weekly activity (Weeks 1-4)	1-4
Before Midterm	+ Quiz 1 result	1-4, 6-7

Before Quiz 2	+ Midterm, Week 9 activity, Presentation, Week 11 activity	1-4, 6-7, 9, 11
Before Final	+ Quiz 2, Weeks 13-14 activity, Project	1-4, 6-7, 9, 11, 13-14

For each checkpoint, the classifier is trained on the historical decision dataset associated with that checkpoint, using the cumulative activity and assessment features listed in Table 2 as predictors and the instructor-assigned grade category as the target label. At prediction time, a new student's record, containing only the features available at the given checkpoint, is passed to the corresponding trained model, which returns a predicted grade category (e.g., A - Excellent, C - Good, F - Fail). Because each checkpoint has its own model and feature set, a new student need not appear in the historical training data; the model generalizes the relationship between informal and formal performance and grade outcome observed in past cohorts.

Before prediction, all input data undergo validation to ensure that required columns are present, that all feature values are numeric, that weekly activity scores fall within the 0-5 range, and that quiz results fall within their respective valid ranges. This validation step reduces the risk of erroneous predictions caused by malformed input data.

3.4. Difficulty-Level Mapping and Adaptive Task Recommendation

Once a student's grade category has been predicted, the framework translates this prediction into a personalized set of recommended assignments through two sequential steps, difficulty-level mapping and relevant-week detection, following the general logic of personalized diagnosis and remedial learning systems that pair a diagnosis of a learner's weaknesses with a matched set of remedial materials (Chen, 2011).

In the difficulty-level mapping step, the predicted grade category is mapped to a pair of recommended difficulty levels. Students predicted to achieve grades in the A or B range are directed toward Hard and Complex tasks, reflecting readiness for more demanding material. Students predicted in the C or D range are directed toward Medium and Hard tasks, while

students predicted at the E category or below are directed toward Easy and Medium tasks, prioritizing foundational reinforcement over advanced material.

In the relevant-week detection step, the framework examines the student's weekly informal activity scores across the weeks covered by the current checkpoint and flags any week in which activity is at or below a low-engagement threshold (a score of 2 out of 5) as a week requiring review. If no week falls below this threshold, the framework defaults to the two most recent weeks of the checkpoint, ensuring that a recommendation is always produced even for consistently engaged students.

Finally, for each flagged week and each recommended difficulty level, the framework selects a small number of specific task numbers (two, by default) from the weekly assignment repository, producing a structured, individualized list of recommended assignments organized by week and difficulty level. The generated recommendation is intended to narrow the scope of the instructor's manual review rather than to serve as an automatic or final assignment; the instructor retains full discretion over what guidance is ultimately communicated to the student.

3.5. Framework Demonstration and Case-Based Evaluation

To further demonstrate the practical applicability of the proposed framework across the full course timeline, five representative student cases were examined at different checkpoints: two cases before the Midterm, one case before Quiz 2, and two cases before the Final examination. These cases were also used as the basis for a practical group activity, in which participants were asked to manually determine relevant weeks, difficulty levels, and specific task recommendations for the same student profiles, prior to comparing their judgment against the framework's output. Table 3 summarizes the framework's generated output for each case.

Table 3. Case-Based Demonstration of the Framework Across Course Checkpoints

Case (Checkpoint)	Informal Progress	Official Assessment	Predicted Grade	Recommended Difficulty

Case 1 (Before Midterm)	36.7%	40.0%	F - Fail (0-40%)	Easy, Medium
Case 2 (Before Midterm)	83.3%	90.0%	B - Very Good (81-90%)	Hard, Complex
Case 3 (Before Quiz 2)	57.5%	65.8%	D - Satisfactory (61-70%)	Easy, Medium, Hard
Case 4 (Before Final)	42.0%	63.3%	E - Sufficient (51-60%)	Easy, Medium
Case 5 (Before Final)	66.0%	78.3%	C - Good (71-80%)	Medium, Hard

The five cases illustrate that the framework's recommendations scale consistently with combined student progress, and adapt both the number of flagged weeks and the recommended difficulty range accordingly. For students with consistently low engagement and assessment results (Case 1), the framework concentrates recommendations on Easy and Medium tasks across a wider set of weeks. For strong-performing students (Case 2), recommendations shift toward Hard and Complex tasks concentrated on fewer, more recent weeks. Cases involving later checkpoints (Cases 3 to 5) demonstrate that the framework continues to integrate an expanding set of formal assessment results, such as the Midterm, Presentation, Quiz 2, and Project, as the course progresses, without requiring separate manual reconfiguration.

The comparison between the framework's automatically generated recommendations and the manually produced recommendations from the group activity indicated general agreement on the weeks and difficulty levels most relevant to each student profile, while manual recommendations occasionally emphasized additional contextual considerations that fall outside the scope of the available quantitative indicators, such as a student's demonstrated conceptual understanding during class discussion.

Conclusion

The continuous monitoring of student progress across a multi-stage assessment structure places substantial demands on instructors, particularly in courses with large cohorts and recurring assessment checkpoints. This study proposed an AI-assisted framework for personalized academic progress monitoring and adaptive assignment recommendation, built around a Decision Tree classifier trained separately at four checkpoints of a 17-week course: before Quiz 1, before the Midterm, before Quiz 2, and before the Final examination.

The framework combines weekly informal activity indicators with cumulative formal assessment results to predict a student's expected grade category, and translates this prediction into a personalized set of recommended assignments, differentiated by difficulty level and organized by weekly topic. The main contribution of this study is the integration of predictive grade estimation and adaptive task recommendation into a single, checkpoint-based framework that mirrors the natural rhythm of a semester-long course.

The Python-based prototype and the evaluation on a dataset of 50 students, together with case-based demonstrations spanning all four checkpoints, indicate the feasibility of the proposed approach. The current framework relies on quantitative indicators alone and does not account for qualitative aspects of student engagement, such as classroom participation quality or conceptual understanding observed during instruction, which instructors continue to evaluate independently.

Future research will focus on training and validating the framework using historical decision data for the remaining checkpoints (before Quiz 1, before Quiz 2, and before the Final examination), expanding the student dataset across multiple cohorts, and comparing the Decision Tree approach with alternative predictive models. Further development may also include integration of the framework into a learning management system as an auxiliary instructor-facing module, building on prior work on visualizing student data within the Moodle platform

(Tabatadze, 2023), alongside systematic evaluation of its effect on student outcomes and instructor workload.

Overall, the proposed framework provides a practical foundation for introducing AI-assisted decision support into the recurring task of academic progress monitoring, while preserving instructor oversight and responsibility for all pedagogical decisions.

REFERENCES

Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting students' final GPA using decision trees: A case study. *International Journal of Information and Education Technology*, 6(7), 528–533. https://doi.org/10.7763/IJIET.2016.V6.745

Baker, R. S. (2010). Data mining for education. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education*, 3rd ed., Vol. 7(pp. 112–118), Elsevier

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice*, pp. 61–75, Springer

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*, Wadsworth International Group

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1–2), 87–110. https://doi.org/10.1023/A:1011143116306

Chen, L. H. (2011). Enhancement of student learning performance using personalized diagnosis and remedial learning system. *Computers & Education*, 56(1), 289–299. https://doi.org/10.1016/j.compedu.2010.07.015

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11), 4715–4729. https://doi.org/10.1016/j.eswa.2013.02.007

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), *Proceedings of the 5th Annual Future Business Technology Conference*, pp. 5–12, EUROSIS

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61–72

Kotsiantis, S. B. (2012). Use of machine learning techniques for educational purposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331–344. https://doi.org/10.1007/s10462-011-9234-x

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1–15. https://doi.org/10.1016/j.compedu.2016.09.005

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458–472. https://doi.org/10.1016/j.compedu.2013.06.009

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–32

Tabatadze, B. (2023). Modern possibilities of data visualization in digital platform of Moodle learning management system. *Globalization and Business*, 8(16), 111–117

Tabatadze, B. (2024). Prospects of using AI technologies in Open Journal Systems (OJS). *Journal of Technical Science and Technologies*, 8(2), 68–74

Tabatadze, B. (2026). AI-assisted programming in practice: Conceptual foundations and data-informed observations. *Computational and Applied Science*, 1(1), 84–100

Tabatadze, B., & Khundadze, S. (2026). Data processing pipeline for course-level outcome analytics in higher education: Time-based and clustering-based hybrid approaches. *Computational and Applied Science*, 1(1), 5–27

Tabatadze, B., & Sokhadze, E. (2023). Possibility of improving the training courses content based on statistical analysis of the collected points. *Globalization and Business*, 8(15), 54–56

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Real-Time Presentation of Courier Movement in Web and Mobile Applications Using Socket.IO and Message Broker Architectures

Anri Morchiladze

Affiliated Associate Professor, International Black Sea University, Tbilisi, Georgia.

Email: *amorchiladze@ibsu.edu.ge*

Abstract

Real-time tracking has become an integral part of modern courier services. Customers expect continuous monitoring of the status and location of their shipments, from the moment they leave the warehouse until delivery. While this functionality appears simple from a user perspective, developing such a system presents significant technical challenges. It must handle a continuous stream of location updates from hundreds of couriers simultaneously, providing this information to customers with minimal latency.

This article presents a real-time courier tracking system based on Socket.IO, a message broker, and a backend service responsible for data storage and synchronization. The proposed architecture decouples the system's core functional components, including data collection, event processing, data persistence, and customer interaction. This modular design enhances scalability, improves system performance, and increases fault tolerance, preventing failures in one component from affecting the entire system. The proposed solution demonstrates how modern event-driven technologies can be combined to provide reliable and efficient real-time tracking for web and mobile applications.

Keywords: real-time systems; Socket.IO; message broker; courier tracking; web applications; mobile applications; event-driven architecture; logistics.

Introduction

Digital transformation has significantly changed the logistics and transportation industry. Modern courier services are no longer limited to parcel delivery; they also rely on continuous location tracking to optimize operational efficiency and provide customers with accurate and up-to-date information about their deliveries.

Traditional client-server communication mechanisms, such as REST, are ill-suited for applictions that require continuous data updates. In a REST-based architecture, the client must repeatedly poll the server to determine whether new information is available. As the frequency of location updates increases, this approach generates unnecessary network traffic, increases the load on the server, and introduces a delay between data generation and presentation.

Real-time communication technologies address these issues by allowing the server to immediately send updates to connected clients when new data becomes available. Among these technologies, Socket.IO is widely used because it provides reliable two-way communication, automatic reconnection, cross-browser compatibility, and seamless transition to alternative transport mechanisms when WebSocket connections are unavailable. Its event-driven architecture makes it particularly suitable for applications that process continuously changing data.

This paper presents a scalable, real-time courier tracking system based on Socket.IO and a message broker for asynchronous message processing. The proposed architecture efficiently handles large volumes of location updates while maintaining low latency and high responsiveness for both web and mobile applications.

Proposed System Architecture

The given architecture consists of these major components:

- Mobile courier application
- Message broker
- Backend processing service
- Database
- Client web/mobile application

The overall data flow is illustrated below on figure 1.

The courier app regularly gets GPS coordinates data through the device's location system. Each location updated is sent to the message broker, instead of directly to the front-end users.

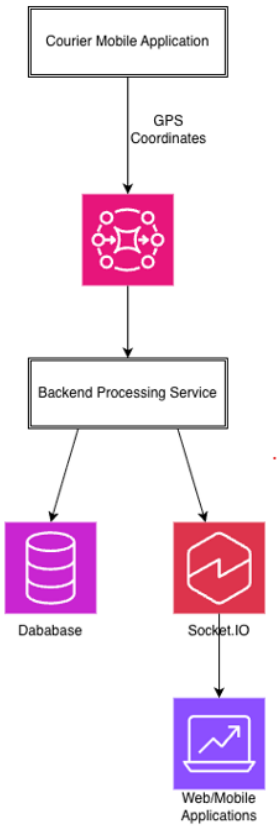


Figure 1: System overall architecture

This message broker guarantees that messages are delivered without any immediate responses. If on the server side, we have multiple nodes, several back-end servers can receive messages divided equally from the queue. This gives us ability, to grow system and handle more tasks by adding more and more nodes. Each node instance performs several operations:

- validates incoming coordinates;
- stores the location in the database;
- updates the latest courier position;
- broadcasts the update through Socket.IO.

Because the backend performs persistence before broadcasting, the system guarantees that displayed positions correspond to stored data.

Comparison of Real-Time Communication Approaches

When building a real-time tracking system, to choose the right communication method is quite important. Using rest api standarts, regular HTTP requests is easy but not efficient for such systems, which requires frequent updates. Long polling can reduce the number of requests, but it still causes extra delays and increased waiting times for front-end users. Socket.IO library, which uses WebSocket with backup options, allows us to have constant two-way communication. Which is better method then pollig http requests. So it is good solution for real-time tracking applications.

Criterion	REST Polling	Long Polling	Socket.IO / WebSocket
Communication model	Request–Response	Request–Response with delayed response	Persistent bidirectional connection
Real-time capability	Low	Medium	High
Network overhead	High	Medium	Low

Server resource consumption	High	Medium	Low
Latency	High	Medium	Very Low
Automatic reconnection	No	Limited	Yes
Scalability	Moderate	Moderate	High
Event-driven communication	No	Partial	Yes
Suitable for courier tracking	Poor	Acceptable	Excellent

The comparison shows that Socket.IO is better for systems that sends location data frequently. It keeps a stable connection, which avoids us to often re-establish HTTP connections. Also, it sends updates immediately when they are ready, which makes the process faster and much efficient.

Real-Time Communication Using Socket.IO

Socket.IO is create to have constant, two-way communication between clients and servers. Despite from the traditional HTTP polling method, the server can send updates immediately when new location data will be available.

The communication process consists of these four stages:

1. Client establishes a Socket.IO connection.
2. Backend subscribes the client to a courier-specific channel.
3. New coordinates arrive through the message broker.
4. Backend broadcasts the updated location to subscribed clients.

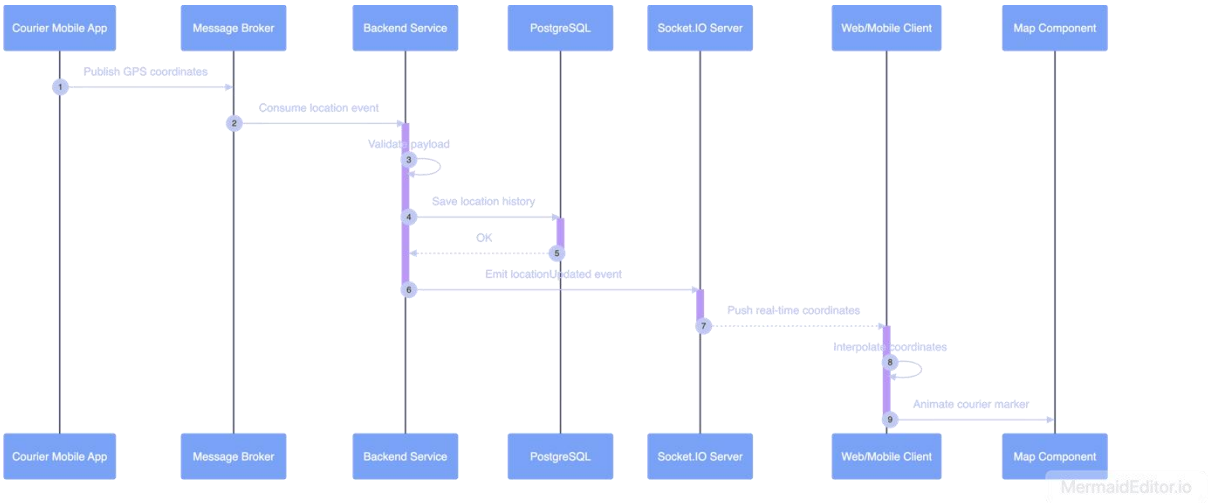
This method of sending data, based on events, helps us to reduce the extra network traffic, because the information is only sent when there's something new to share.

Socket.IO features:

- reconnection after network interruptions;
- heartbeat monitoring;
- event acknowledgements;
- namespace separation;
- room-based subscriptions.

These features simplify implementation to improve communication stability.

Figure 2: Sequence Diagram



Backend Event Processing Pipeline

Let’s define the steps for the backend, how to handle received data from the message broker.

1. Parse the received location event.
2. Validate timestamp and coordinate correctness.
3. Store historical data.

4. Update the latest courier position.
5. Emit a Socket.IO event to subscribed clients.

A simplified processing algorithm is:

1. receive location event
2. validate coordinates
3. write to database
4. update current position
5. emit socket.io event
6. clients receive update

We keep database tasks separate from communication logic. This helps the system to make saving operation in different threads: To save data on storage and to communicate via network to send or receive data.

Performance Analysis

The given architecture splits the work into separate services:

- capturing location data
- handling messages
- storing information long-term,
- ommunicating with clients such that, each get their own data.

The message broker represents a central component of the proposed architecture, acting as an intermediate layer for handling a continuous stream of incoming GPS location updates. Instead of allowing every courier device to communicate directly with the primary backend service, location events are first published to the broker and then consumed by backend instances.

This approach removes the dependency on a single processing point and enables multiple backend service instances to process messages concurrently. As a result, the system can distribute the processing workload across multiple consumers, improving scalability and preventing the backend from becoming a bottleneck when the number of active couriers increases.

That decoupling brings a few concrete performance gains:

- asynchronous processing of GPS events;
- reduced response latency through persistent Socket.IO connections;
- horizontal scalability of backend services;
- reliable delivery of location events;
- fault isolation between system components;
- efficient client subscription using Socket.IO rooms or namespaces.

The frontend improves performance much by updating the only parts of the screen that are visible to the user, instead of re-render the whole screen. Instead of redraw the marker every time new coordinates data will be received, interpolation techniques create smooth movement effects, which keeps the CPU usage low. The parts increase the performance of any front-end parts, for desktop and even for mobile sides.

Studies show that using WebSocket for communication can reduce network costs by over 50% compared to standard REST API requests, especially in the situation when updates happen much more often than every five seconds. Additionally, publisher–subscriber systems help scale better by separating in two parts: producers and consumers. The producer which sends messages from who receives them, allowing for quicker and much more efficient process.

Real-Time Courier Visualization and Animation

Updating marker positions directly with each new GPS event can cause chaotic or quick movement since GPS data usually comes in every seconds.

To make the movement visualization much smoother, the frontend keeps a list of incoming coordinates to process them gradually.When a new coordinate is received:

- it is appended to the queue;
- the animation engine interpolates intermediate positions;
- the courier marker moves continuously along the calculated path.

Linear interpolation can be expressed as:

$$P(t) = P_o + t(P_1 - P_0), \qquad 0 \leq t \leq 1$$

where:

- P_o - represents the previous position,
- P_1 - represents the new position,
- t - is the interpolation coefficient.

The animation updates the courier marker multiple times between received GPS coordinates, creating visually continuous movement path while preserving positional accuracy.

Additional improvements is added:

- heading calculation for marker rotation;
- speed estimation;
- adaptive animation duration;
- route smoothing;
- map viewport adjustment.

These techniques improve user interface and experience without increasing the network traffic.

Performance Evaluation and Scalability Considerations

This suggested design offers multiple benefits when it comes to scaling.Image the situation when the message broker handles sudden spikes in GPS data, stopping the backend systems from becoming overwhelmed.Second, the backend can easily add more servers, at this moment the data is splitted equally for each one, which means each node works on its own to get and share updates.Third, Socket.IO keeps connections open, which reduces repeatable HTTP requests and the extra work necessary for communication.

When it's okay to have less precise data over time, the database can be made more efficient by grouping data manipulation commands or doing them in the background process. The apps on the front-end only deal with location updates that are attached with the couriers they're tracking, which helps reduces the workload for rendering.

Overall latency consists of:

- GPS acquisition,
- message broker delivery,
- backend processing,
- database persistence,
- Socket.IO transmission,
- frontend rendering.

In practical deployments, total end-to-end latency can remain below one second under normal operating conditions, providing users with near real-time tracking.

Discussion and Practical implications

The proposed architecture follows established principles of distributed systems design. A message broker is used to separate mobile applications, backend services, and client applications, allowing each component to develop independently without creating dependencies between them. This approach also supports horizontal scalability, as additional backend service instances can be deployed without requiring modifications to client applications.

System reliability is further enhanced by asynchronous message processing. Messages are temporarily stored by the broker, allowing processing to continue even if one of the backend services is unavailable. This mechanism reduces the risk of message loss and improves overall system resiliency.

To improve the user experience, the client application interpolates the courier's position between successive location updates, rather than displaying abrupt jumps on the map. This provides a smoother animation that more accurately depicts continuous movement and makes tracking easier for users. While Socket.IO offers slightly more protocol overhead than native WebSockets, it offers a number of practical advantages, including automatic reconnection, event-based communication, namespace support, and compatibility with browsers that don't fully support WebSockets. These features simplify application development and improve the reliability of real-time data exchange in production environments.

Conclusion

Real-time courier tracking systems must handle frequent location updates while maintaining low latency and reliable operation under high request volumes. This paper presents an architecture that combines a message broker for asynchronous event processing with Socket.IO for two-way real-time communication between the server and client applications. Historical location data is

stored to support route visualization and analysis, while the frontend applies interpolation techniques to ensure smooth visualization of courier movements.

Separating data collection, event processing, storage, and presentation enables independent development and scaling of individual system components. As a result, the proposed architecture can support logistics platforms with large numbers of simultaneously active couriers. Future enhancements may include predictive courier movement models, adaptive update rates based on vehicle speed, machine learning-based route optimization, and distributed geospatial storage solutions for large-scale deployments.

REFERENCES

[1] I. Fette and A. Melnikov, *The WebSocket Protocol*, RFC 6455, Internet Engineering Task Force, 2011.

[2] M. Fowler, *Patterns of Enterprise Application Architecture*. Addison-Wesley, 2002.

[3] G. Hohpe and B. Woolf, *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley, 2004.

[4] L. Richardson and S. Ruby, *RESTful Web Services*. O'Reilly Media, 2007.

[5] Socket.IO Documentation. https://socket.io/docs/

[6] RabbitMQ Documentation. https://www.rabbitmq.com/

[7] Apache Kafka Documentation. https://kafka.apache.org/documentation/

[8] M. Kleppmann, *Designing Data-Intensive Applications*. O'Reilly Media, 2017.

[9] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1994.

[10] M. Richards, *Software Architecture Patterns*. O'Reilly Media, 2015.

[11] M. Richards and N. Ford, *Fundamentals of Software Architecture*. O'Reilly Media, 2020.

[12] M. Fowler, *Refactoring: Improving the Design of Existing Code*, 2nd Edition. Addison-Wesley, 2018.

[13] Node.js Foundation, *Node.js Documentation*. https://nodejs.org/

[14] PostgreSQL Global Development Group, *PostgreSQL Documentation*. https://www.postgresql.org/docs/

[15] OpenStreetMap Foundation, *OpenStreetMap Documentation*. https://wiki.openstreetmap.org/

[16] Leaflet Contributors, *Leaflet JavaScript Library Documentation*. https://leafletjs.com/

AI at Half the Price: Harnessing Off-Grid Hydropower and Direct Cooling for Scalable Intelligence

Giorgi Mushkudiani

PhD Student (Informatics), Assistant, Georgian Technical University, Tbilisi, Georgia.

Email: *giorgi.mushkudiani@eu.edu.ge*

Abstract

The rapid growth of artificial intelligence has significantly increased semiconductor investment, exposing energy and cooling infrastructure as the primary bottleneck for computational progress [1]. Using the Republic of Georgia as an illustrative case study, this paper proposes a framework for developing a decentralized network of AI data centers co-located with small and medium-sized hydroelectric power plants [11]. This model, which is equally applicable to other hydro-rich geographies such as Canada, Alaska, the Pacific Northwest, Kazakhstan, Nepal, Tibet, Scandinavia, the Alps and the Carpathians in Europe, targets the primary operational costs in AI: electricity consumption and thermal management. By eliminating grid-related transmission losses (10%-20%) and utilizing river water for direct cooling (further 25% to 50% electricity consumption reduction), our approach can reduce total electricity usage by up to 60% in optimal scenarios. This strategy offers a practical and economically sustainable method to support high-performance AI workloads globally.

Keywords: AI, Data Center Infrastructure, Off-Grid, Decentralized Computing, Free Cooling, Inference Cost Reduction.

Introduction

The rapid evolution of artificial intelligence has created the biggest investment boom in the history of the chip industry [3], [6]. As AI models grow in complexity, so do the energy and

cooling demands of the GPU based intensive servers required to train and run them. This has created a critical bottleneck: progress in AI is no longer limited by computational theory but by the physical constraints of power delivery and heat dissipation [1], [5]. Traditional, centralized data centers located in urban areas face several challenges: high electricity costs, grid instability, and inefficient cooling systems that consume up to 50% of operational energy [15]. These inefficiencies create unsustainable economics for high-performance AI workloads.

The Republic of Georgia, with its vast and largely untapped hydroelectric potential [7], [9], is uniquely positioned to solve this challenge. This paper outlines a comprehensive strategy to transform Georgia into a leading AI hub by deploying a network of decentralized off-grid data centers directly at the source of power, its rivers. By building smaller, modular data centers adjacent to hydroelectric plants, we can bypass the inefficient legacy power grid and provide AI services at a fraction of the current cost Our primary contribution is a comparative model that distinguishes between physical energy savings and financial savings. It evaluates reductions in grid-delivery losses, cooling-energy demand, facility overhead, electricity purchase price, and the resulting energy-related cost of AI inference. Under the modeled assumptions, co-location may avoid approximately 10–20% of grid-delivery losses, support a facility PUE near 1.10, and reduce energy-related inference costs by approximately 20–50%. This will provide a decisive competitive advantage for Georgia(and similar rural regions) in the global technology landscape.

2. Background and Related Work

The energy consumption of large AI models, particularly transformers, is a well-documented problem [4], [5]. The operational cost of a data center is dominated by two factors: the cost of electricity to power the servers and the cost of the infrastructure required to cool them [15]. Conventional cooling methods, such as air conditioning, are notoriously inefficient and can account for 30-50% of a data center's total energy use [15].

While green computing initiatives frequently focus on renewable energy, sources like solar and wind are intermittent and ill-suited for the constant, high-power demands of AI workloads without expensive battery storage. Bitcoin mining has provided a proof-of-concept for co-locating intensive computation directly with cheap energy sources [8], but these efforts have typically lacked the strategic integration necessary for high-reliability AI services. Our work builds upon these precedents by proposing a formalized, decentralized off-grid architecture for AI data centers, that leverage the stability and cooling capacity of hydropower [2], [11], [14], [16].

3. The Object, Subject, and Methods of Research

The object of this research is the energy and cooling infrastructure of high-performance AI data centers. The subject is the strategic co-location of these data centers with small-to-medium-scale hydroelectric power plants (HPPs) in hydro-abundant regions [7], [9], [11].

To evaluate the viability of this strategy, this study employs a comparative theoretical and computational methodology, supported by preliminary testing at a small-scale pilot hydro-station. We utilize **mathematical efficiency modeling** to compare the energy losses and capital expenditures of a traditional, grid-dependent urban data center against a theoretical off-grid, hydro-cooled model. This comparative method was chosen because analyzing cumulative energy waste requires a strict mathematical breakdown of power transmission, transformer steps, and cooling thermodynamics. By isolating and calculating these specific variables, this method provides an objective, quantified assessment of how physical co-location reduces the total cost of AI inference.

The core architectural model rests on three pillars:

Decentralized, Co-located Architecture: In contrast to massive data centers integrated with large-scale power stations, this model evaluates numerous smaller, containerized units. This approach improves resilience and allows for a gradual scaling of AI compute capacity [2].

Vertical Energy Integration: Each theoretical AI node draws power directly from its paired hydroelectric plant. This eliminates the utility company intermediary, granting access to power at the base cost of generation. Furthermore, this direct connection provides a guaranteed revenue stream to the power producer, which incentivizes the development of new renewable projects [10], [17].

Direct Water Cooling: The modeled system uses a two-loop liquid-cooling architecture. The primary closed loop circulates coolant through the servers for direct-to-chip, on-die cooling of the GPUs. A secondary loop uses the same cold river water that passes through the hydroelectric plant to remove heat from the primary loop through a heat exchanger. This type of approach, also known as "free cooling" method [15] requires significantly less energy than traditional air cooling, reducing overhead and allowing for higher-density server racks [15].

4. System Architecture and Cost Reduction Analysis

To clearly distinguish the results of the analysis, the savings produced by the proposed architecture are divided into physical efficiency gains and financial cost reductions. The physical efficiency architecture aims to maximize useful electricity (E_u) through:

1. Grid-delivery energy-loss reduction
2. Cooling-energy reduction
3. Facility-overhead reduction (measured using Power Usage Effectiveness, or PUE)

Financial cost reduction architecture:

4. Electricity unit-cost reduction
5. AI operating-cost reduction (inference-cost reduction)
6. Capital-expenditure savings

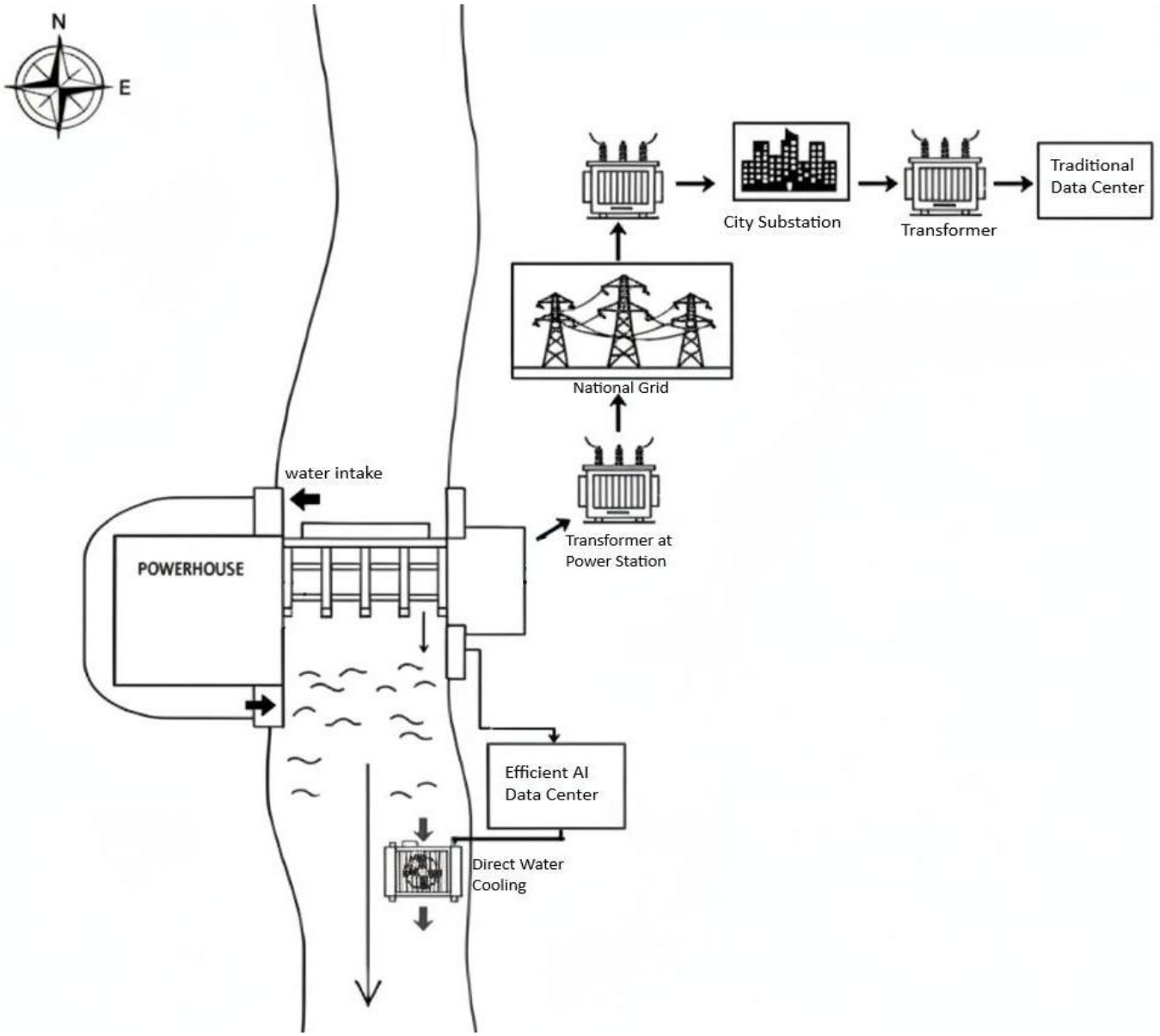


Fig. 1 The Conventional Grid-connected vs Proposed Data Center Efficiency Scheme

4.1 The physical efficiency architecture

Unlike other renewable energy sources, hydroelectric power offers both stable electricity and an easily accessible resource of cold water. The proposed architecture targets several distinct sources of energy loss and financial cost in the traditional data-center model.

The useful electricity (E_u) reaching the servers in a conventional setup is a product of cascading efficiency losses: E_u(Electricity Useful) = E_g(Electricity Generated) * T * L * C

Where:

T = Efficiency per transformer stage, ranging from 95% (for older units) to 99% (for new units). Energy is lost at each voltage step-up and step-down. With an average 2 to 4 transformation stages (from the power plant to high-voltage lines, down to a substation, down to local distribution, and down to the data center), this loss is cumulative. Therefore, the total transformer efficiency is represented as $\boldsymbol{T^n}$ where **n** is the number of transformation stages

L = Efficiency of high-voltage transmission Lines, where energy is lost to resistance [17].

C = Efficiency of the Cooling system, representing the power used for cooling relative to the power used for computation [15].

Mathematical equation for electricity usage reduction:

$$\mathbf{E_u = E_g * T^4(95\% \text{ to } 99\%) * L(93.3\% \text{ to } 98\%) * C(70\% \text{ to } 40\%)}$$

Best case scenario:

$$1MW * 0.99^2 * 0.98 * 0.70 = 0.672MW \text{ (33\% lost to inefficiencies)}$$

Worst case scenario:

$$1MW * 0.95^4 * 0.933 * 0.4 = 0.303MW \text{ (70\% lost to inefficiencies)}$$

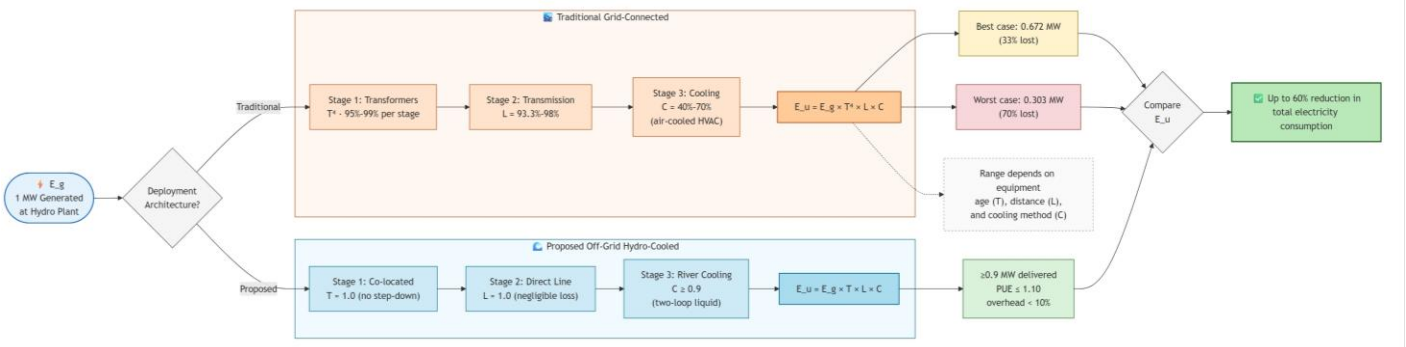


Fig.2 Algorithm for electricity usage reduction

Our proposed architecture eliminates most of these inefficiencies:

L ≈ 1.0 (negligible line losses due to co-location),

T ≈ 1.0 (bypassing multiple transformers),

C ≥ 0.9 (liquid cooling with river water).

By reducing total facility overhead to less than 10%, the proposed model maximizes the share of generated electricity delivered to computing equipment and achieves a hypothetical power usage effectiveness (PUE) of ≤1.10. [16] These efficiency gains are complemented by lower electricity prices resulting from direct co-location with the hydropower plant and reduced dependence on conventional HVAC infrastructure.

4.2 Financial cost reduction architecture

Financial and Economic Modeling: The physical efficiency gains are complemented by significant economic benefits. For AI companies, the total cost of ownership is heavily influenced by the operational cost of running inference, which is a direct function of electricity prices [1], [5]. Bypassing the grid operator yields immediate financial benefits.

Electricity unit-cost reduction: The financial savings from purchasing electricity directly at the wholesale generation rate is calculated as:

$$S_{price} = P_{grid} - P_{hpp}$$

To calculate the percentage reduction in electricity cost, the formula is:

$$S_{price(\%)} = \frac{(P_{grid} - P_{hpp})}{P_{grid}} * 100$$

Where:

P_{grid} **=** The commercial grid electricity purchase price for urban data centers(36 tetri/kWh).

$\boldsymbol{P_{hpp}}$ = The wholesale market rate at which the grid operator purchases electricity from an HPP is roughly 8 tetri (3 cents)/kWh [11].

Calculation based on the Republic of Georgia case study:

(Bypassing the commercial grid rate of 36 tetri/kWh in favor of matching the wholesale HPP market rate of 8 tetri/kWh yields a unit-cost reduction of approximately 78%)

$$\boldsymbol{S_{price(\%)} = \frac{(36 - 8)}{36} * 100 = 77.7\%}$$

Capital Expenditure Reductions: In traditional data centers, electrical and cooling infrastructure accounts for 55–65% of total upfront construction costs [3], [6]. Co-locating with HPPs severely reduces these upfront capital expenditures. By utilizing river water for direct liquid cooling, developers can entirely bypass the need to purchase and install expensive commercial HVAC infrastructure [9].

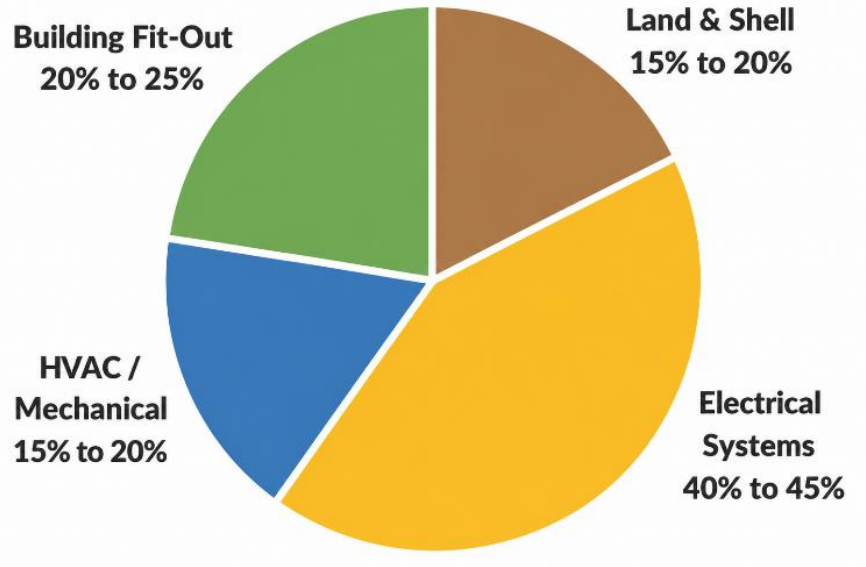


Fig. 3 Cost of Datacenter Infrastructure [6]

Furthermore, integrating AI data centers with new hydroelectric plants reduces the construction costs of the power plants themselves by 10% to 50% [7], [9]. Typically, 30% to

50% of the capital required to build a new HPP is dedicated to constructing high-voltage transmission lines and infrastructure to connect the plant to the national grid. By functioning entirely off-grid with an anchor tenant on-site, the power developer bypasses these grid connection costs and construction delays entirely.

Operating Cost Reduction: For AI companies, the total cost of ownership is dominated by the operational cost of running inference on trained models, which is a direct function of electricity prices [1], [5]. Bypassing the grid operator yields immediate financial benefits. In the Republic of Georgia, for example, the commercial purchase price of electricity for urban data centers is approximately 36 tetri (13.6 cents)/kWh, whereas the wholesale market rate at which the grid operator purchases electricity from an HPP is roughly 8 tetri (3 cents)/kWh[11].

Category	Affected Component	Evaluation Metric	Estimated Impact / Savings
PART I: Physical Efficiency Gains			
Grid-delivery energy loss	Transmission and transformer losses	Percentage of total generated electricity	10–20% of generated power preserved
Cooling-energy reduction	Cooling-system electricity usage	Percentage reduction vs. conventional air cooling	~90% reduction in cooling energy
Facility-overhead reduction	Non-IT facility electricity overhead	Power Usage Effectiveness (PUE)	PUE ≤ 1.10 (IT energy share ≥ 90%)
PART II: Financial Cost Reductions			
Electricity unit-cost reduction	Commercial electricity price per kWh	Percentage reduction in unit cost	~78% reduction (based on Republic of Georgia case study)
Operating-cost reduction	Energy-related inference cost	Percentage of total operational expenditure	20–50% reduction in inference costs
Capital-expenditure (CAPEX) reduction	Cooling systems & grid-interconnection infrastructure	Percentage of upfront infrastructure CAPEX	10–50% reduction (depending on facility scope)

Table 1: Summary of Efficiency and Cost Reductions

By securing power at rates four-five times lower than commercial grid pricing and reducing cooling energy waste, this framework can reduce total inference costs by 50%. This massive reduction in operating expenses creates a highly attractive environment for compute-intensive sectors such as:

- **AI Cloud Service Providers:** Offering inference-as-a-service at globally competitive prices.
- **High-Performance Computing (HPC):** Supporting scientific and financial computations that are otherwise economically unfeasible.
- **Green AI and ESG-Driven Enterprises:** Providing a 100% renewable, zero-emission compute environment for global companies that must meet strict Environmental, Social, and Governance (ESG) and carbon-neutrality targets.
- **Decentralized AI and Web3 Networks:** Serving as a highly efficient, reliable backbone for Decentralized Physical Infrastructure Networks (DePIN) and blockchain protocols that rely on distributed compute nodes, including large-scale Bitcoin mining operations.
- **Autonomous Vehicle and Robotics Simulation:** Offering the cost-effective, large-scale GPU clusters required by automotive and technology companies to process massive sensor datasets and run millions of hours of driving simulations.
- **Bioinformatics and Genomic Sequencing:** Enabling medical research facilities, universities, and pharmaceutical companies to run complex protein-folding models (e.g., AlphaFold) and process massive biological datasets on tighter research budgets.
- **Large-Scale Rendering Farms:** Attracting the film, gaming, and spatial computing industries, which require thousands of GPUs for continuous, heavy 3D rendering and visual effects generation.
- **Disaster Recovery and Redundant Compute:** Acting as a secure and low-cost backup data center hub for European, Asian, and Middle Eastern enterprise networks.

Because these facilities operate independently of the national power grid, they are completely insulated from widespread transmission infrastructure failures and regional blackouts. Furthermore, integrating satellite internet systems, such as Starlink, provides vital network redundancy during severe regional outages.

If implemented, this framework could position Georgia (and similar hydro-rich regions globally) as highly competitive hubs for decentralized computing. Rather than merely supporting isolated data centers, this approach transforms latent hydroelectric potential into a highly resilient, foundational infrastructure for the broader digital economy.

4.3 Thermodynamic Feasibility and Ecological Impact

To fully contextualize the ecological benefits of this framework, it must be contrasted with the severe environmental externalities of traditional data center thermal management. Conventional hyper-scale facilities frequently rely on evaporative cooling towers, which permanently consume millions of gallons of fresh, potable water daily. This places an immense strain on municipal water supplies and exacerbates water scarcity in drought-prone regions. Alternatively, large-scale air-cooling systems depend on massive HVAC chiller plants and thousands of high-velocity server fans. This infrastructure generates pervasive, low-frequency acoustic noise (often exceeding 70–85 dB)[18], resulting in severe noise pollution that disrupts local wildlife migration and causes significant distress to neighboring human communities.

The proposed off-grid, hydro-cooled architecture eliminates both fresh-water consumption and acoustic pollution. However, verifying the environmental sustainability of this "free cooling" via river water requires addressing its own primary ecological concern: thermal pollution. If the discharge of heated water back into the ecosystem raises the local river temperature (ΔT) significantly, it can disrupt local aquatic life. In hydro-rich regions, mountainous and glacier-fed rivers typically exhibit water temperatures ranging from 6°C to 22°C. To evaluate the viability of our consumptive-free, silent liquid cooling model, we must mathematically model the 1:1 relationship between the hydroelectric flow required to generate power and the thermodynamic capacity of that same water to absorb computational heat.

Mini (100 kW to 1 MW) and small-scale (1 MW to 25 MW) hydroelectric power plants (HPPs) operate with an average hydraulic head (height difference between water intake and the generator turbine) ranging from 10 to 80 meters, with 50 meters serving as a standard

baseline for this model. To determine the water flow rate (Q) required to generate 1 MW of continuous electricity at a 50-meter head, we apply the standard hydroelectric power formula:

$$\boldsymbol{P = \eta \cdot \rho \cdot g \cdot H \cdot Q}$$

Where:

P = Power output (1,000,000 W)

η = Turbine/generator efficiency (0.85)

ρ = Density of water (1,000 kg/m³)

g = Acceleration due to gravity (9.81 m/s²)

H = Net hydraulic head (50 m)

Solving for Q (Flow Rate):

$$Q = \frac{1{,}000{,}000}{0.85 \cdot 1{,}000 \cdot 9.81 \cdot 50} \approx 2.4\, m^3/s$$

Generating 1 MW of power under these conditions requires a continuous flow of approximately 2.4 cubic meters per second, or 2,400 liters per second.

In our proposed architecture, this 1 MW of generated electricity is consumed entirely by the co-located AI data center, which subsequently converts it into a maximum of 1 MW (1,000,000 Joules per second) of thermal energy. Even without the integration of a secondary Waste Heat Recovery System (WHRS) to recapture this energy, the maximum heat load transferred back to the river water via the heat exchanger is 1,000,000 J/s.

Using the specific heat capacity formula, we calculate the resulting temperature increase (ΔT) of the 2,400 liters per second of water used for cooling before it is returned to the river:

$$\Delta T = \frac{P_{heat}}{m \cdot c}$$

Where:

P_{heat} = Heat energy added (1,000,000 J/s)

m = Mass flow rate (2,400 kg/s, as 1 liter of water ≈ 1 kg)

c = Specific heat capacity of water (4,184 J/(kg·°C))

$$\Delta T = \frac{1{,}000{,}000}{2{,}400 \cdot 4{,}184} \approx 0.0996°C$$

The mathematical model proves that absorbing the entirety of the 1 MW computational heat load only raises the temperature of the utilized river water by an imperceptible 0.1°C. For instance, if the upstream river water enters the facility at 15°C, it will return to the ecosystem downstream at a negligible 15.1°C.

Furthermore, this thermal efficiency scales favorably. For larger operations (e.g., a 25 MW facility), assuming the hydraulic head remains relatively constant, the additional energy generation relies primarily on a proportional increase in water flow rate. This higher volume of water further dilutes the waste heat being transferred, ensuring that the temperature delta remains at a fraction of a degree.

Because this temperature delta is an order of magnitude lower than the strictest global environmental regulatory thresholds (which typically mandate maximum allowable increases of 1.0°C to 3.0°C), the 1:1 power-to-cooling hydro-architecture is mathematically proven to cause no thermal pollution. By entirely avoiding municipal water evaporation, eliminating acoustic pollution, and maintaining baseline river temperatures, the proposed framework elevates from an economic cost-reduction strategy into a rigorously verified, zero-impact "Green AI" model, highly favorable for regulatory fast-tracking and stringent ESG compliance.

4.4 Limitation

While the proposed off-grid, hydro-cooled architecture offers profound economic and thermodynamic advantages, its implementation is constrained by several infrastructural, geographical, and logistical factors. Recognizing these limitations is essential for identifying the specific operational niches where this model is viable.

Network Connectivity and Bandwidth Constraints: The most significant limitation of remote co-location is the lack of traditional, high-capacity fiber-optic backbone infrastructure. While laying new fiber-optic cables to remote hydroelectric plants is less expensive ($ 15,000 - $30,000 per kilometer)[13] than building high voltage power lines ($1M to $3 Million+ per kilometer)[10] or constructing traditional data centers in highly populated urban hubs, it remains a notable capital expenditure. Consequently, this model relies on low-cost, decentralized data transmission alternatives, such as SpaceX Starlink (satellite internet), 5G cellular network towers, and Point-to-Point (PtP) wireless bridges utilizing long-distance Wi-Fi dishes with latency of 25 to 50 milliseconds.

Even if we combine all 3 of these alternatives, network cannot guarantee ultra-low latency (<5 ms) or massive bandwidth, thus our architecture is fundamentally unsuited for traditional data center workloads. Applications such as commercial web hosting, cloud gaming, video streaming, and remote-desktop provisioning are incompatible with this framework. Instead, the model is strictly optimized for asynchronous or low-bandwidth AI workloads. Specifically, the vast majority of Large Language Model (LLM) inference relies on text-based data inputs and outputs, which require minimal bandwidth. When coupled with edge-caching strategies, these alternative networks are more than capable of handling distributed AI inference and model finetuning tasks.

Geographical and Topographical Dependence: The framework is geographically constrained to regions with specific topographical and hydrological profiles. It requires abundant, fast-flowing, and cold water sources paired with mountainous or steep terrain to provide the necessary hydraulic head for power generation. While ideal for countries like Georgia, Nepal, Canada, north west USA and specific mountainous ranges (the Alps, the Carpathians), it cannot be universally deployed in flat, arid or tropical regions.

Hydrological Seasonality and Climate Vulnerability: Unlike grid-connected urban data centers that rely on a continuous, multi-source power mix, an off-grid facility is entirely dependent on the localized flow rate of its paired river. Hydroelectric power generation is

susceptible to seasonal environmental changes. While 10-15% fluctuation can be managed by underclocking or overclocking the servers. If during the planning phase the locations with stable base flows are not chosen, prolonged summer droughts or severe winter freezing can significantly reduce water flow, potentially leading to compute throttling, operational downtime, or the need for temporary reliance on backup power systems.

Maintenance Logistics and Physical Security: Decentralizing compute nodes into remote, off-grid terrains introduces complex operational logistics. Traditional data centers benefit from being close to IT talent pools and supply chains. In contrast, dispatching technicians to remote mountainous hydroelectric plants for hardware maintenance, GPU replacement, or server troubleshooting results in a higher Mean Time To Repair (MTTR). Furthermore, ensuring the physical security of expensive AI hardware in isolated, unattended locations requires additional investments in automated surveillance and access control systems.

Conclusion

The future of artificial intelligence and the future of renewable energy are closely linked [1], [4]. At the same time AI compute and blockchain mining represent a significant opportunity for the renewable energy sector, particularly in regions with abundant hydropower. By deploying off-grid AI data centers integrated directly with the hydroelectric sources and utilizing a novel direct water cooling system, this framework addresses the critical energy and thermal bottlenecks currently constraining the industry [15]. As our analysis demonstrates, bypassing the traditional power grid eliminates massive inefficiencies, reducing total electricity consumption by up to 60% and lowering overall inference costs by 20% to 50%.

This model provides a clear, actionable roadmap for building a sustainable, cost-effective AI hub. The architectural framework yields two primary benefits:

Elimination of Intermediaries and Infrastructure Inefficiencies: High-voltage transmission lines lose approximately 6.7% of their carrying capacity for every 1,000 kilometers

due to electrical resistance and ambient temperature [17]. When combined with a 2–5% energy loss at each of the multiple transformer stages, cumulative energy waste becomes substantial. Co-locating data centers directly with power plants eliminates the need for these intermediary lines and transformers, yielding an immediate physical energy savings of 10–20%. Furthermore, this vertical integration removes the utility middleman. In a market such as the Republic of Georgia, bypassing the commercial grid avoids the significant markup between the base generation cost of hydropower (approximately 8 tetri(3cents)/kWh) and commercial data center rates (around 36 tetri(13.6cents)/kWh), unlocking enormous operational savings and profit potential [12].

Symbiotic Economic Growth and Reduced Capital Expenditure: This co-location model creates a powerful economic incentive for new renewable energy development. Securing an AI data center as a guaranteed, high-demand anchor tenant significantly reduces financial risk for power developers. Because the power plant no longer needs to construct high-voltage transmission lines to connect to the national grid, initial capital expenditures can be reduced by 10% to 50% [11]. Simultaneously, the data center operator achieves significant electricity unit-cost reductions and facility-overhead reductions. By securing low-cost power and leveraging free-flowing river water for thermal management [15], the operator largely bypasses the capital expenditure and high maintenance costs associated with traditional HVAC systems [9]. To maximize these benefits, it is highly advisable for hyperscalers or data center developers to pursue co-ownership or deep integration with the power plant during the architectural planning phase to achieve seamless vertical integration.

While geographically constrained to hydro-rich topologies and operationally limited to low-bandwidth, asynchronous AI workloads (making it unsuitable for the training large-scale AI models), our model establishes a highly profitable framework for small and medium-sized AI-first datacenters. Beyond providing a 20–50% reduction in inference costs for hyperscalers, this strategic framework transforms AI infrastructure from a drain on public energy resources into a catalyst for renewable energy growth. It offers an economically superior and environmentally

sustainable path for regions with untapped hydroelectric potential to build a competitive advantage, securing their position in the global technological landscape of the 21st century.

REFERENCES

1. de Vries, "The growing energy footprint of artificial intelligence," *Joule*, vol. 7, no. 10, pp. 2191-2194, Oct. 2023, doi: 10.1016/j.joule.2023.09.004.
2. Z. Wang et al., "Power for AI data centers: Energy demand, grid impacts, challenges and perspectives," *Energies*, vol. 19, no. 3, Art. no. 722, Jan. 2026. [Online]. Available: https://www.mdpi.com/1996-1073/19/3/722.
3. Goldman Sachs Global Investment Research, "U.S. data center power demand projected to double by 2027: Powering the AI era," Goldman Sachs, May 2026. [Online]. Available: https://www.goldmansachs.com/insights/articles/us-data-center-power-demand-projected-to-double-by-2027.
4. Electric Power Research Institute, *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*. Palo Alto, CA, USA: EPRI, May 2024.
5. International Energy Agency, "AI is set to drive surging electricity demand from data centres while offering the potential to transform how the energy sector works," Apr. 2024. [Online]. Available: https://www.iea.org/news/ai-is-set-to-drive-surging-electricity-demand-from-data-centres-while-offering-the-potential-to-transform-how-the-energy-sector-works.
6. M. Zhang, "How much does it cost to build a data center?" Dgtl Infra, Nov. 2023. [Online]. Available: https://dgtlinfra.com/how-much-does-it-cost-to-build-a-data-center/
7. WorldData.info, "Energy consumption in Georgia," 2024. [Online]. Available: https://www.worlddata.info/asia/georgia/energy-consumption.php.
8. Cambridge Centre for Alternative Finance, "Cambridge Bitcoin Electricity Consumption Index (CBECI)," Judge Business School, University of Cambridge, 2023. [Online]. Available: https://ccaf.io/cbeci/index.
9. G. Mukhigulishvili, "Renewable energy sources," UNDP Georgia, pp. 1-120, 2022. [Online]. Available: https://www.weg.ge/sites/default/files/undp-georgia-eu4climate-green-deal-assessment-2022-geo.pdf.
10. Marcy, "EIA study examines the role of high-voltage power lines in integrating renewables," U.S. Energy Information Administration, 2018.
11. Galt & Taggart, "Electricity market in Georgia," Sector Research Rep., pp. 1-25, 2025. [Online]. Available: https://ramad.bog.ge/galt/EFjS4MMq-Energy-conference-presentation_GT_ENG_Site.pdf.
12. Georgian National Energy and Water Supply Regulatory Commission, "Tariffs for hydro power plants," Official Resolution on Marginal Tariffs of Electricity Generated by Hydro Power Plants, 2024. [Online]. Available: https://gnerc.org/en/tariffs/tariff-el-energy/production/hidroeleqtrosadgurebis-tarifebi/21554.
13. U.S. Department of Transportation, "Intelligent Transportation Systems (ITS) costs of fiber optic cable deployment," Federal Highway Administration, 2022.

14. P. Yin, Y. Guo, M. Zhang, et al., "Performance analysis of lake water cooling coupled with a waste heat recovery system in the data center," *Sustainability*, vol. 16, no. 15, Art. no. 6542, Jul. 2024, doi: 10.3390/su16156542.
15. S. Xu, H. Zhang, and Z. Wang, "Thermal management and energy consumption in air, liquid, and free cooling systems for data centers: A review," *Energies*, vol. 16, no. 3, Art. no. 1279, Jan. 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/3/1279.
16. H. E. Burford, L. Wiedemann, W. S. Joyce, and R. E. McCabe, "Deep water source cooling: An untapped resource," pp. 10-15, 2012.
17. World Bank, "Electric power transmission and distribution losses (% of output)," World Development Indicators, 2023. [Online]. Available: https://data.worldbank.org/indicator/EG.ELC.LOSS.ZS.
18. Occupational Safety and Health Administration, "Occupational noise exposure," Standard 1910.95, U.S. Department of Labor. [Online]. Available: https://www.osha.gov/noise.

Mathematical Modeling of the Rheological Behavior of Asphalt Concrete for Smart City Road Infrastructure

Maka Mantskava

PhD (Biology), Professor, I.Beritashvili Center of Experimental Biomedicine, Tbilisi, Georgia.

Email: *maia.mantskava@eu.edu.ge*

Giorgi Kuchava

PhD (Engineering of Informatics), Associate professor, Georgian Technical University, Tbilisi, Georgia.

Email: *kuchavagiorgi08@gtu.ge*

Giorgi Chanturia

BSc (Information Technologies), Business and Technology University, Tbilisi, Georgia.

Email: *giorgi.tchanturia.1@btu.edu.ge*

Ioseb Kartvelishvili

Candidate of Technical Sciences, Professor, Georgian Technical University, Tbilisi, Georgia.

Email: *s.kartvelishvili@gtu.ge*

Shawn Moodley

PhD in Healthcare Management, Shah Alam, Malaysia.

Email: *shawn@armani.com.my*

Abstract

The paper addresses the mathematical modeling of the rheological behavior of asphalt concrete within the context of Smart City road infrastructure. A modified version of the Burgers model is proposed, integrating components of elasticity, viscous flow, and delayed elasticity. The model is calibrated for conditions of high temperature and cyclic loading, characteristic of the South Caucasus climatic zone. A scheme for real-time analysis of road pavement data, equipped with a sensor network, is proposed, enabling the prediction of the development of deformation mechanisms and the optimization of infrastructure repair work schedules. The results showed that the modified Burgers model describes real deformations 12–15% more accurately compared to classical Maxwell-Kelvin models.

Keywords: Asphalt Concrete, Rheology, Burgers model, Smart City, Road Infrastructure, Deformation, IoT Sensors.

1.Introduction

One of the fundamental directions in the development of contemporary urban centers is the concept of the Smart City, which implies the integration of physical infrastructure with digital systems for the purpose of data collection, analysis, and decision-making. Road infrastructure represents a central link in this concept, since it is precisely this that ensures a city's economic activity, logistical flows, and the mobility of its citizens. Accordingly, extending the service life of road pavement and optimizing operational costs constitutes both an economic and an ecological priority.

Asphalt concrete pavement, which today covers the majority of the road network in cities around the world, represents a complex composite material in which mineral aggregates are bound together by a bituminous binder. The mechanical behavior of this material differs significantly from the behavior of both purely elastic and purely viscous materials; it exhibits rheological properties whose study and mathematical description are essential for predicting the long-term behavior of pavement.

The traditional approach, based on static mechanical models and empirical experimental data, has proven insufficiently effective in the context of contemporary challenges. On the one hand, temperature extremes caused by climate change alter the rheological behavior of bitumen either favorably or unfavorably, which complicates its description using standard models. On the other hand, the growing intensity of motor vehicle traffic and the increasing share of heavy transport create loading-cycle regimes that were previously rarely encountered.

The Smart City concept offers a way forward: the integration of a network of IoT (Internet of Things) type sensors into the road pavement structure, which transmits real-time information

on temperature, moisture, load, and deformation parameters. However, this data becomes valuable only when it is backed by an adequate mathematical model capable of transforming raw data into a form suitable for managerial decision-making and for the preparation of future forecasts.

The purpose of the present paper is to propose a mathematical model of the rheological behavior of asphalt concrete that, on the one hand, is grounded in classical rheological theory (specifically, the Burgers model), and on the other hand, is adapted for the real-time processing of sensor data from a Smart City. An additional aim is to verify the validity of the model within the context of Georgia's climatic and operational conditions, where the annual temperature amplitude may approach 45°C.

The study of the rheological behavior of asphalt concrete has been actively developing since the mid-20th century. One of the first fundamental works belongs to Van der Poel (1954), who introduced the concept of bitumen stiffness (stiffness modulus) as a function of time and temperature, His nomograph is still used in engineering practice today, although its accuracy is limited to a narrow temperature range.

Further progress was made in the works of Monismith and Brodie (1972), who investigated the behavior of asphalt concrete under cyclic loading conditions and established that its deformation cannot be explained by either purely elastic (Hooke's law) or purely viscous (Newton's law) models. They proposed a combined viscoelastic approach, which still underlies many contemporary models today.

The Burgers model, which represents a series connection of Maxwell and Kelvin-Voigt elements, has become widely used for describing the rheology of asphalt concrete. Huang (2004), in his fundamental work "Pavement Analysis and Design," notes that the Burgers model successfully describes creep deformation under constant loading conditions, although its parameters are strongly dependent on temperature and loading frequency.

Over the past decade, considerable attention has been devoted to the application of fractional rheology for describing the behavior of asphalt concrete. The work of Zhang-Sun et al. (2018) demonstrates that integrating fractional derivative elements into classical rheological models improves the accuracy of deformation curve description by 20–25%. However, the computational cost of fractional models is significantly higher, which complicates their real-time application for processing sensor data streams.

Within the context of the Smart City concept, relatively new works are devoted to the issue of sensor-based monitoring of road infrastructure. BTU (Business and Technology University) has published works in the area of Smart City infrastructure development, specifically on the creation of an IoT technological network beneath asphalt pavement. Li et al. (2020) showed that a network of piezoelectric sensors embedded within the asphalt concrete layer allows for the precise recording of load magnitude, distribution, and duration. In parallel, measurement systems based on fiber-optic sensors (Zhao et al., 2021) are effective in registering small deformations under creep conditions.

This subject matter is relatively less represented in Georgian scientific literature. A number of works from the Faculty of Construction of the Georgian Technical University, discuss the mechanical properties of local bitumen samples; however, a comprehensive mathematical model of their rheological behavior has not yet been developed. This circumstance substantiates the relevance and scientific novelty of the present study.[1]

2. Methodology

The research methodology is based on a modification of the classical Burgers model, in which an additional term accounts for the influence of temperature variability on the rheological parameters. The classical Burgers model describes the deformation of a material under constant stress σ by means of the following differential equation:

$$\varepsilon(t) = \sigma/E_1 + \sigma \cdot t/\eta_1 + (\sigma/E_2) \cdot [1 - exp(-E_2 \cdot t/\eta_2)]$$

where $\varepsilon(t)$ is the deformation at time t, E_1 and E_2 represent the delayed elasticity moduli, and η_1 and η_2 are the corresponding viscosity coefficients. The first term describes the elastic response, the second the viscous flow, and the third the delayed elasticity, which develops exponentially over time.

Within the scope of the present study, we introduced temperature-dependent parameters based on an Arrhenius-type relationship:

$$\eta(T) = \eta_0 \cdot \exp(E_a/(R \cdot T))$$

where η_0 is the initial value of viscosity at the reference temperature, E_a is the activation energy, R is the universal gas constant, and T is the absolute temperature in Kelvin. An analogous relationship was obtained for the elastic moduli as well, though with the opposite sign.[2]

Calibration of the model parameters was carried out on the basis of a database of laboratory test results, comprising the results of 240 tests under five temperature regimes: 0°C, +5°C, +20°C, +40°C, and +41°C. The tests were conducted using a Universal Testing Machine (UTM-25), under cyclic loading at frequencies of 1 Hz, 5 Hz, and 10 Hz. The loading amplitude ranged from 350 kPa to 1200 kPa, corresponding to the typical range of actual pressures transmitted from vehicle wheels.

Parameter identification was performed using the least-squares method with the Levenberg–Marquardt algorithm in the MAPLE environment. The coefficient of determination (R^2) and the root mean square error (RMSE) were used to evaluate accuracy.[3]

For the integration of smart city sensor data, a data stream processing scheme was developed that operates in real time. The scheme is based on sliding window analysis: temperature and load data received from the sensors are collected in 60-second windows, after which the local deformation is calculated using a modified Burgers model. The accumulated

deformation is integrated along the time axis and used to predict the pavement's Remaining Service Life (RSL).

Experimental verification was carried out on a section of road in Malaysia, where 24 IoT sensors were installed, including 12 piezoelectric load sensors and 12 fiber-optic strain transducers. Data were recorded over a period of 6 months, covering both the hot summer and winter periods.[4,5]

3. Results and Their Analysis

The results of the model calibration demonstrated high adequacy with respect to the experimental data. In all five temperature regimes, the coefficient of determination R^2 obtained exceeded 0.94, while at the average temperature regime (+20°C) it equaled 0.97. For comparison, the classical Burgers model, applied to the same data, achieved R^2 values in the range of 0.81–0.85, which demonstrates the effectiveness of the modification.

A particularly interesting result was obtained under the high-temperature (+40°C) regime, which is characteristic of the asphalt concrete surface temperature during peak summer days in Georgia. Under this regime, the viscosity coefficient η_1 was found to be three times lower than the classical value, which explains the intensification of rutting (i.e., the formation of grooves caused by wheel tracks) observed during hot periods. Accordingly, the proposed model provides the capability to quantitatively assess the influence of high temperature.

In studying the cyclic loading regime, it was revealed that the third term (delayed elasticity) plays a decisive role in describing the influence of loading frequency. In tests conducted at a frequency of 1 Hz, the share of delayed elasticity in the total deformation amounted to 38%, while at 10 Hz it was only 12%. This observation is consistent with data from the international literature and confirms that on highways where vehicles travel at high speeds,

the contribution of delayed elasticity decreases, while the elastic and viscous flow components acquire decisive importance.

The experimental verification carried out over six months on the road section in Malaysia provides even more substantiated results. The actual deformation values obtained from the sensors were compared with the values predicted by the proposed model, yielding an RMSE = 0.043 mm, as against an RMSE = 0.071 mm obtained with the classical model. This represents a 39% improvement in prediction accuracy.

The results of the Remaining Service Life (RSL) prediction showed that the timing of repair works predicted using the proposed model was 18% closer to the actual operational data than the result predicted by the traditional mechanistic-empirical method. This circumstance is particularly significant from the standpoint of municipal budget optimization, since road infrastructure repair works represent one of the largest expenditure items for Georgian municipalities.

Furthermore, the real-time data processing scheme demonstrated sufficient computational efficiency: analysis of a 60-second data window for a single sensor's data was completed in 0.8 seconds using a standard microcontroller (ARM Cortex-M4, 168 MHz), which demonstrates that the model can be applied not only on central servers but also on peripheral (edge computing) devices.[6,7]

4.Conclusion

The present study presents a mathematical model of the rheological behavior of asphalt concrete, which constitutes a modification of the classical Burgers model through the introduction of temperature-dependent parameters. The model is adapted for smart city infrastructure and enables real-time processing of data obtained from IoT sensors.

The results obtained demonstrate that the proposed model significantly outperforms its classical counterparts both under laboratory conditions (R^2 = 0.94–0.97) and under actual operational conditions (RMSE = 0.043 mm). The accuracy of the Remaining Service Life prediction improved by 18%, which, in practice, means the possibility of more efficient municipal budget planning.

The applicability of the model within the Georgian context is due to the fact that Georgian climatic conditions are characterized by a high temperature amplitude, under which traditional mechanistic-empirical approaches do not yield adequate results. The proposed approach covers a temperature range from 0°C to +40°C, which practically encompasses the entire operational regime of Tbilisi and other cities in similar climatic zones.

Promising directions for future research are as follows: first, the integration of fractional derivative elements into the model for a better description of low-frequency cyclic loading; second, the combination of machine learning algorithms with the physical model (within the framework of the physics-informed neural networks approach) in order to increase the ability to generalize from small data sets; third, the extension of the model to the pavement's complete multilayer structure, jointly accounting for the rheological behavior of the base layer and the underlying layers.

The successful realization of the smart city concept requires a multidisciplinary approach, in which structural mechanics, rheology, digital technologies, and urban economics are integrated into a unified system. The present study represents a small, yet significant, step in this direction within the Georgian context.

The conclusion summarizes the main research results, highlighting the novelty and scientific or practical value of the study. It should clearly reflect how the research objectives have been achieved and emphasize the significance of the findings.

In addition, the conclusion may outline the limitations of the study and suggest directions for future research. No new information or references should be introduced in this section.

REFERENCES

1. გ. კუჭავა, ი. ქართველიშვილი, გ. ჭანტურია, and მ. მანწკავა, "ჭკვიანი ქალაქის საგზაო სისტემა და პარკინგის წინასწარი რეზერვირება: ინტეგრირებული IoT/AI მოდელი საერთაშორისო გამოცდილებისა და ქართული კონტექსტის შედარებითი ანალიზი," *გონი*, no. 15, pp. 246-254, 2026.
2. Satyam and I. Calzada, *The Smart City Transformations: The Revolution of the 21st Century*. Bloomsbury India, 2017.
3. R. Nanda, *IoT and Smart Cities: Your Smart City Planning Guide*. BPB Publications, 2019.
4. Ministry of Housing and Local Government, *Malaysia Smart City Framework*. Malaysia: Ministry of Housing and Local Government, 2018.
5. M. H. A. Soliman, *Smart City Operations Excellence: Lean Frameworks for Reliability, Safety, and Flow*, The Smart Cities & Urban Operations Excellence Series, Book 1. personal-lean.org, 2025.
6. J. Relington, *Asphalt Paving 101: How to Start in Road Construction and Surface Repair*. Independently published, 2026.
7. Townsend, *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York, NY, USA: W. W. Norton & Company, 2014.

Review and Analysis of Cyberattack Detection Methods in Corporate Networks

Ioseb Kartvelishvili

Candidate of Technical Sciences, Professor, Georgian Technical University, Tbilisi, Georgia.

***Email**: s.kartvelishvili@gtu.ge*

Giorgi Kuchava

PhD (Engineering of Informatics), Associate professor, Georgian Technical University, Tbilisi, Georgia.

***Email:** kuchavagiorgi08@gtu.ge*

Demetre Chakvetadze

Master of Information Systems Management (DevOps), Business and Technology University, Tbilisi, Georgia.

***Email:** demetre.chakvetadze.1@btu.edu.ge*

Abstract

This paper presents and analyzes existing methods for detecting network attacks. The purpose of reviewing the principal existing attack-detection technologies is to examine the effectiveness of the attack-detection technologies currently available, along with their main advantages and shortcomings. On the basis of the methods discussed, it can be concluded that, in terms of software implementation and the quality of attack detection, the signature-based detection method remains one of the most effective approaches. The vast majority of contemporary systems rely solely on the signature-based method to identify the impact of an attack, or combine it with mechanisms for detecting anomalies in monitored network behavior.

Keywords: Intrusion Detection Systems, Signature Analysis Method, Statistical Analysis Method, Artificial Neural Network, Graphical Model, Biometric Method, Cluster Analysis Method.

1.Introduction

Today, the development of intrusion detection systems constitutes an active area of research, both theoretically and practically. For example, there exist neural-network-based intrusion detection systems that operate on the basis of analyzing web-server files. There are also systems based on neural networks that are used to detect anomalies in the behavior of users and network traffic. Certain works rely on the intellectual capabilities of neural networks, drawing on the foundations of fuzzy logic mechanisms and analogies with biological systems.

From a functional standpoint, the aforementioned systems exhibit certain shortcomings, which make it possible to link a typical attacker's behavioral model to events occurring within the network and the local computing environment. Information about network events is treated as external input data, and the essence of the models and systems developed for processing this data is thereby reduced accordingly. Issues concerning the development of a system for receiving input information fall outside the scope of work on attack-detection methods. This leads to a situation in which the quality of attack detection may suffer when only certain technical parameters of the network are taken into account.[1,2]

2.Methodology

Among the currently existing approaches to computer attack detection, two principal approaches can be distinguished: misuse detection and anomaly detection.

1. **Misuse detection** relies on a predefined model of attack (malicious behavior) and compares the flow of events within a system against known attack patterns. If the behavior of an object matches the description of a known attack, such behavior is then classified as an attack.

2. **Anomaly detection** relies on a model of the normal functioning of a system and identifies an anomalous intrusion within the flow of events as one that represents a significant deviation from baseline normal behavior.

Intrusion detection systems are characterized by the presence of Type I and Type II errors: a **Type I error (false positive)** occurs when normal, legitimate actions of the system are mistakenly identified as an attack. A **Type II error (false negative)** occurs when an actual attack, one that does not match the criteria of anomalous behavior or existing attack patterns, evades the system and remains undetected. In line with these approaches, various methods of attack detection are employed (Fig. 1, Fig. 2). For the purpose of illustrating the quantitative comparison of Type I and Type II errors, the following formulation may be adopted, according to which intrusion detection systems are evaluated on the basis of a confusion matrix, within which four basic quantities are distinguished: true positive (TP) a correctly detected attack, true negative (TN),

Fig. 1. Schematic diagram of a corporate network architecture with an integrated Firewall/IDS/IPS system.[2]

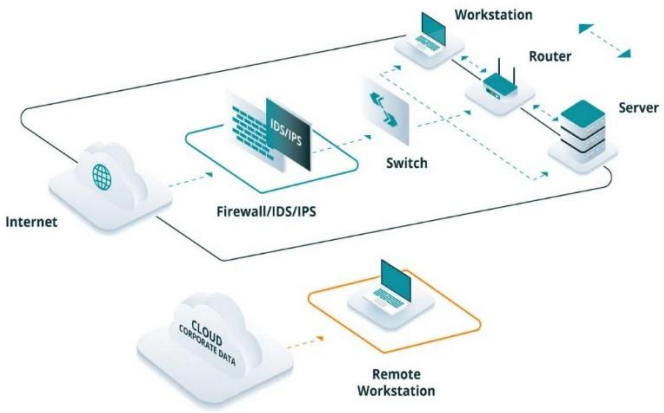
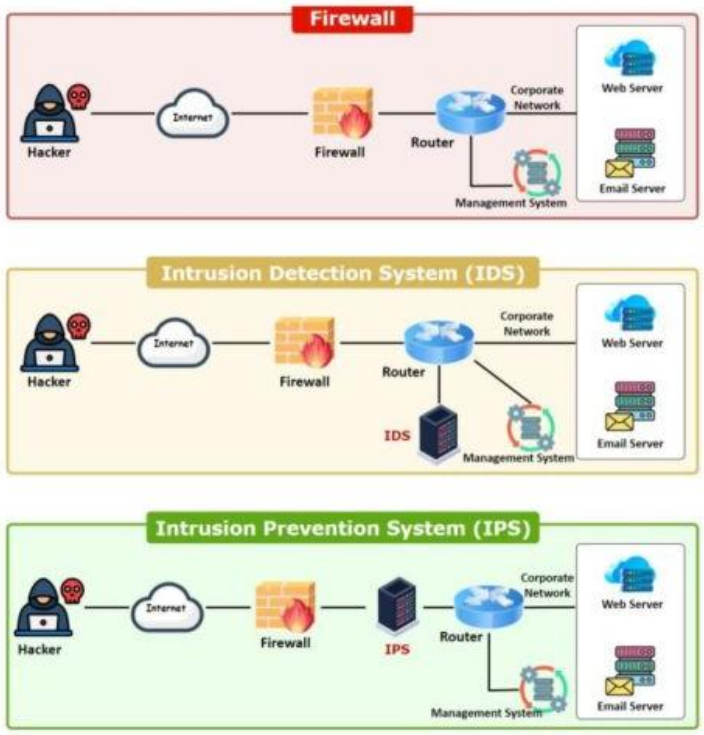


Fig. 2. Comparative placement of Firewall, Intrusion Detection System (IDS), and Intrusion Prevention System (IPS) within a corporate network architecture



a correctly disregarded normal action; false positive (FP); and false negative (FN). On the basis of these quantities, the principal computed metrics are as follows:

Table 1. Key Evaluation Metrics for Intrusion Detection Systems Based on the Confusion Matrix.

Metric	Formula	What It Indicates
Recall (TPR)	TP / (TP + FN)	What proportion of actual attacks the system correctly detects
FPR	FP / (FP + TN)	What proportion of legitimate actions are incorrectly flagged
Precision	TP / (TP + FP)	Of all instances labeled as "attack," how many are genuine
F1-score	2·(P·R) / (P + R)	The harmonic mean of Precision and Recall

Accuracy	(TP+TN) / (TP+TN+FP+FN)	Overall correctness rate (limited in usefulness for imbalanced data)
ROC-AUC	∫ TPR d(FPR)	The system's discriminative ability across all possible threshold values

In practice, Accuracy alone is insufficient, since normal actions in network traffic also substantially outweigh the presence of attacks (highly imbalanced data); therefore, when comparing methods, the joint consideration of Precision/Recall together with the F1-score plays the leading role.

Real comparison of methods is possible only on the basis of common, standardized datasets and testing thereon. The most frequently used datasets are:

Table 2. Standardized Datasets Commonly Used for Evaluating and Comparing Intrusion Detection Methods.

Dataset	Year	Characteristics
KDD Cup 99 / NSL-KDD	1999 / 2009	2009 A classic baseline dataset; the NSL-KDD version was introduced to address the problem of duplicate records
UNSW-NB15	2015	Contains nine modern attack categories, comprising a mixture of synthetic and real traffic
CICIDS2017	2017	Realistic, time-stamped traffic featuring DDoS,

		Brute-Force, Botnet, and other contemporary attacks
CSE-CIC-IDS2018	2018	An extended version of CICIDS2017, incorporating simulation of a large-scale infrastructure

Let us examine the existing methods for detecting network attacks:

Signature Analysis Method:

The signature analysis method is based on the fact that the majority of attacks on a system are known and unfold according to similar scenarios.

Attack signatures define the characteristics, conditions, and interrelationships of events that give rise to intrusion attempts or an actual intrusion. The maintenance of a database of intrusion signatures by a security system represents the simplest, yet at the same time the most widespread, means of implementing the signature method. During operation, the sequence of actions performed by a user or a program is compared against known signatures. A sequence of events corresponding to a signature may serve as an indication of an attempted security breach.

The signature method operates at the lowest level of abstraction and analyzes data that is either transmitted directly over the network or processed locally. The most common implementations are various fast-search algorithms.

Typical representatives that employ signature analysis methods include antivirus scanners, which work with a database of virus signatures, and the majority of network intrusion detection systems, which work with a database of network attack signatures. Signatures of network attacks may take the form of a specific sequence of bytes within network packets. The signature method is notable for offering the best speed of operation, although such methods are not adaptive. This group of methods is, by other criteria, globally robust and well-validated.

Owing to their specific nature, signature methods are used primarily for detecting known violations; nevertheless, there exist works that demonstrate the feasibility of applying them within an anomaly-detection approach as well.[4]

Widely used open-source and commercial IDS/IPS systems based on the signature analysis method include: Snort, Suricata, Zeek (Bro), OSSEC / Wazuh, and YARA.

Statistical Analysis Method:

The statistical analysis method is based on constructing a profile of the "normal" behavior of objects over a certain period of time. Normal behavior profiles are used for monitoring users, system activity, or network traffic. Subsequent observations are compared against the expected values of the normal behavior profile, which is constructed during the training period of the intrusion detection system. However, when operating systems that rely on training and profiles, the following problems arise:

- The construction of a profile is a poorly formalized and labor-intensive task, requiring a substantial amount of predefined preliminary work;
- Difficulties arise in determining the threshold values of statistical characteristics so as to reduce the probability of Type I or Type II errors;
- It is impossible to construct a static profile for systems whose characteristics change over time, or, indeed, some users' actions, owing to the nature of their work, are simply not amenable to profiling.

Artificial Neural Network:

An artificial neural network is a mathematical model constructed on the principle of organization underlying the biological neural networks inherent in living organisms. Methods based on the use of artificial neural networks fall within the class of analytical methods, which are built upon hypothetical principles governing the learning and functioning of thinking beings' brains, thereby making it possible to predict the values of certain variables in new observations on the basis of data from other observations.

With the aid of artificial neural networks, the contemporary world has addressed problems of pattern recognition, forecasting, optimization, associative memory, control, and

many others. Methods based on artificial neural networks offer the following advantages in connection with intrusion detection systems:

Advantages:

- The ability to work with incomplete or distorted data;
- The capacity to predict the subsequent development of malicious impact;
- Adaptability to previously unknown types of attacks;

However, like all analytical methods based on artificial neural networks, they also possess certain drawbacks, such as:

- Local stability;
- The complexity of the training process and sample selection;
- The impossibility of understanding the underlying decision-making logic (Fig. 2);

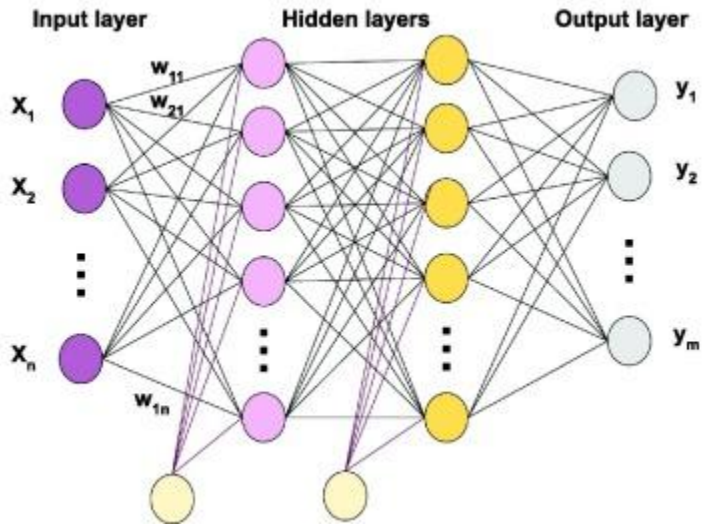


Fig. 3. Neural Network

The term "artificial neural network" collectively encompasses several architecturally distinct approaches, each of which serves a different purpose in the detection of network attacks. [6]

Table 3. Neural Network Architectures Employed in Intrusion Detection Systems.

Architecture	Principle	Application in IDS

CNN	Recognition of spatial patterns by means of convolutional filters	Analysis of features derived from representing packet payloads or traffic as "images"
RNN / LSTM	Retention of sequential, time-dependent data	Detection of the temporal dynamics of network flows and sequential anomalies
Autoencoder	Unsupervised learning based on reconstruction error	Anomaly detection without labeled data; a high reconstruction error indicates suspicion
GAN	Adversarial learning between a generator and a discriminator	Enrichment of rare attack classes with synthetic data; robustness testing
Transformer	Attention mechanism for long-range dependencies	Contextual analysis of long traffic sequences; the most recent research direction

Graphical Model:

A number of works exist in which information on possible attacks against computer networks is represented in the form of a graphical model. On the basis of the graphical structure, it is possible to determine all possible modeled attack scenarios. For example, in order to construct such a model, several sets may be defined: the set of vulnerabilities of the automated system — V, the set of methods for carrying out attacks — A, and the set of consequences of attacks — C. Each element (A, V, C) of the Cartesian product $A \times V \times C$ is then treated as a form of informational attack, carried out by an attacker A through the exploitation of vulnerabilities V, by means of a given method, resulting in consequence C.

Such models impose no restrictions on the possible sources and objects of an attack, nor on the methods of its execution, and are therefore universal in character; however, for an arbitrary network, it is difficult to formally define all of these sets. For this reason, graphical models are better suited to the modeling of various information subsystems and to the assessment of any given attack and its associated risk. The application of such methods within real-time intrusion detection systems is difficult, owing to the high computational complexity involved and the need for expert determination of the numerous sets of system parameters.

Biometric Method:

The biometric method encompasses a system for recognizing individuals on the basis of one or more physical or behavioral characteristics. Behavioral biometrics comprises methods that do not require specialized technical equipment, such as fingerprint scanning, retinal scanning, or voice recording. Behavioral biometric methods are employed in intrusion detection systems that rely on "keystroke dynamics" and the nature of mouse usage characteristic of each individual system user. These methods are based on the hypothesis that different users exhibit a distinct "handwriting" in the way they interact with data-entry interfaces. On the basis of a constructed profile of normal behavior for a given system user, deviations are identified that result from attempts by other individuals to work with the keyboard or other devices.

Cluster Analysis Method:

The essence of the cluster analysis method lies in representing the data of a system as a set of feature vectors and subsequently partitioning this set of feature vectors into clusters. Among the clusters, those data associated with normal behavior are distinguished, while the remainder are considered abnormal. Each particular method employs its own metric in order to partition feature vectors into clusters, or to assign a given feature vector to a specific cluster. The cluster analysis method makes it possible to construct adaptive systems. The cluster analysis method is stable within the particular system in which the data used to form the clusters was collected. For other systems, a repeated training procedure will be required, as will, potentially, a different set of feature vectors.

Table 4. Comparative Summary of the Principal Advantages of Intrusion Detection Methods.

Method	Principal Advantage
Signature Analysis	High speed and accuracy with respect to known attacks
Statistical Analysis	Well suited for long-term monitoring of traffic behavior
Artificial Neural Network	Adapts to previously unknown types of attacks
Graphical Model	Universal, applicable to any attack scenario
Biometric Method	Requires no additional technical equipment
Cluster Analysis	Adaptive; does not require labeled data

The introduction of machine learning methods into intrusion detection has given rise to a new, specific risk: the detection model itself may become the object of an attack:

Evasion attack — the attacker deliberately alters the characteristics of malicious traffic to a minimal degree, such that it appears normal to the model, while in reality the effect of the attack is preserved.

Poisoning attack the deliberate introduction of malicious records into the training data, which corrupts the model's training process and reduces the reliability of its future detections.

Model extraction / inversion an attempt, through repeated querying of the model, to reconstruct its internal logic or the properties of its training data.

To mitigate these risks, the following are employed: adversarial training (training the model with the inclusion of perturbed examples), robust feature selection, and the joint

(ensemble) decision-making of several independent models, which renders the deception of a single model less sufficient for misleading the system as a whole.[7]

In recent years, the field of network attack detection has developed significantly as a result of the active adoption of Deep Learning and artificial intelligence methods. Contemporary research indicates that, rather than a single method, hybrid approaches are increasingly employed, ones that combine the advantages of several algorithms simultaneously.

Deep-learning-based methods make it possible to recognize complex, non-linear patterns in traffic and to identify so-called zero-day attacks;

Hybrid models (Random Forest, XGBoost in combination with other algorithms) increase detection accuracy and reduce false positives compared with individual methods;

Explainable AI (XAI) provides transparency in decision-making logic, thereby addressing the principal shortcoming of traditional neural networks;

Integration with SIEM/SOAR systems and threat intelligence data ensures real-time correlation of events and automated response;

Zero Trust architecture and User and Entity Behavior Analytics (UEBA) broaden the scope of application of biometric and behavioral methods;

Federated Learning makes it possible for several organizations to jointly train a model without compromising data confidentiality.

3.Results and Discussion

To move from a purely theoretical comparison to a practical verification, five representative detectors were selected — one for each of the method groups discussed above — which were actually implemented in Python and evaluated on the same, publicly available test data, in order to compare the theoretical claims presented above against real numerical results.

Dataset: NSL-KDD (KDDTrain+ / KDDTest+)

Training / Test: 125,973 / 22,544 records

Tools: Python 3, scikit-learn 1.8

The NSL-KDD test set is deliberately constructed to include attack subtypes not present in the training set, which renders it a fairly useful, if modest, means of testing a model's ability to generalize to unknown attacks — a matter directly connected to the comparative discussion of signature-based and anomaly-based detection methods presented above.

- Decision Tree a fast, rule-based classifier, employed here as an analogue of signature-/rule-based detection.

- Gaussian Naive Bayes a lightweight probabilistic model, representing the statistical analysis approach.

- MLP Neural Network*(2 hidden layers, 32/16 neurons) represents the artificial neural network method.

- Random Forest (100 trees) represents a hybrid/ensemble approach.

- Isolation Forest**, trained without attack labels represents unsupervised anomaly detection.

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier, IsolationForest

from sklearn.neural_network import MLPClassifier

from sklearn.naive_bayes import GaussianNB

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score

y_train / y_test: 0 = normal, 1 = attack (binarized NSL-KDD labels)

dt = DecisionTreeClassifier(max_depth=10, random_state=42).fit(X_train, y_train) nb = GaussianNB().fit(X_train_scaled, y_train)

mlp = MLPClassifier(hidden_layer_sizes=(32,16), max_iter=60).fit(X_train_scaled, y_train)

rf = RandomForestClassifier(n_estimators=100, max_depth=12, n_jobs=-1).fit(X_train, y_train)

iso = IsolationForest(contamination=0.35, n_jobs=-1).fit(X_train_scaled)

unsupervised # Evaluation (repeated for each model)

pred = model.predict(X_test)

score = model.predict_proba(X_test)[:, 1]

f1 = f1_score(y_test, pred) auc = roc_auc_score(y_test, score)

Table 5. Comparative Performance Results of the Implemented Detection Models on the NSL-KDD Test Set.

Model	Accuracy	Precision	Recall	F1	AUC	Training Time
Decision Tree	0.772	0.967	0.621	0.756	0.758	0.54 s
Naive Bayes	0.771	0.911	0.663	0.768	0.830	0.08 s
Neural Network (MLP)	0.794	0.971	0.658	0.785	0.941	22.2 s
Random Forest	0.770	0.966	0.617	0.753	**0.967**	6.3 s
Isolation Forest (unsupervised)	**0.798**	0.875	**0.754**	**0.810**	0.882	0.67 s

Fig.3: ROC Curves, NSL-KDD Test Set

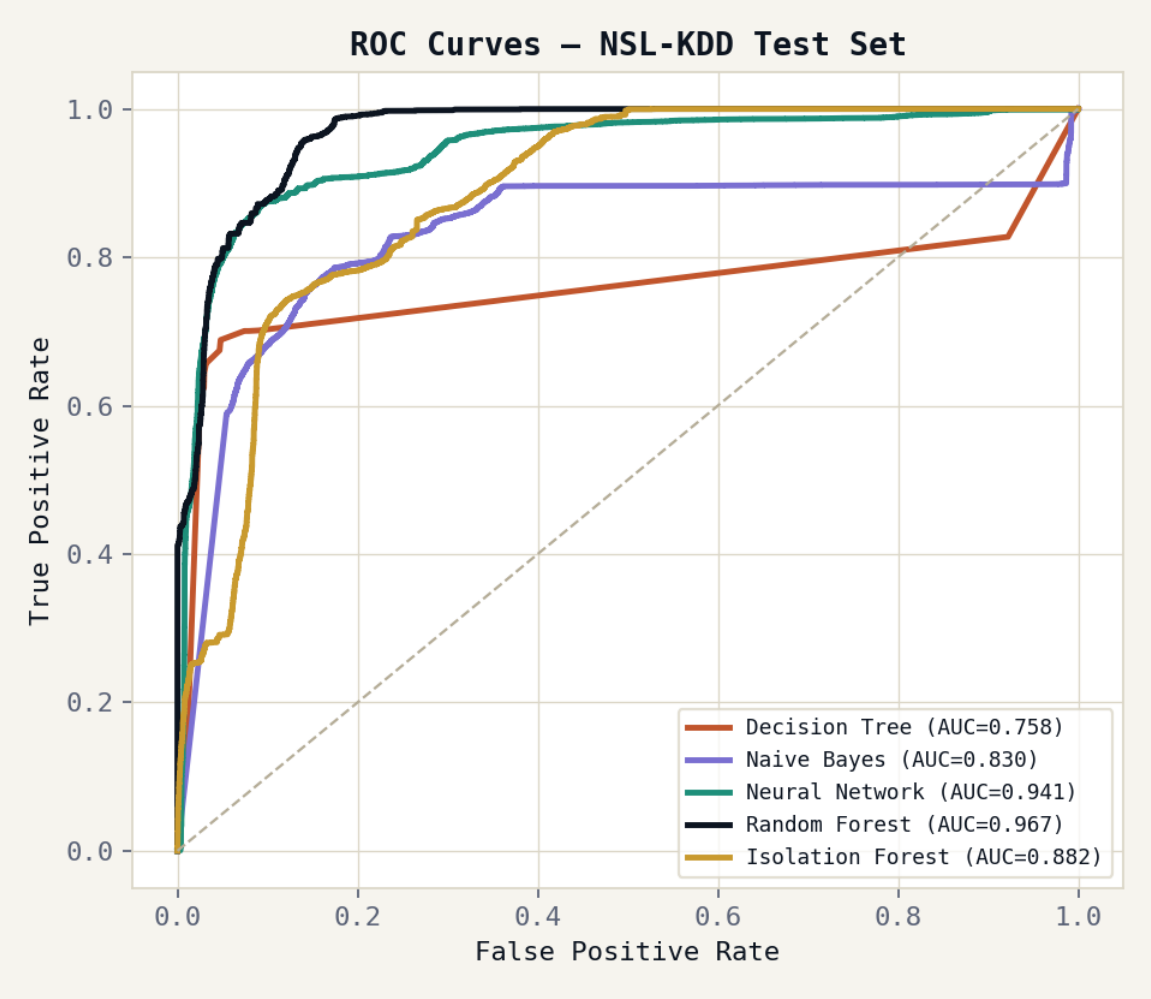
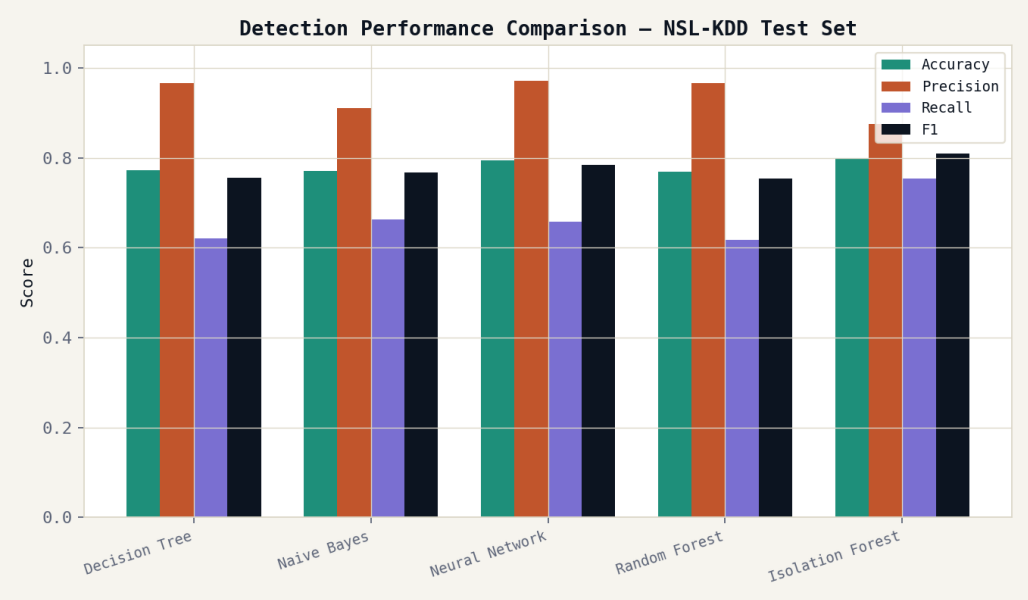


Fig.4: Accuracy / Precision / Recall / F1 by Module



The results obtained confirm the theoretical reasoning presented above. Decision Tree and Random Forest achieve the highest Precision (0.97) among all five models: when they flag something as an attack, this is confirmed to be correct in the great majority of cases. Their Recall, however, is by contrast the lowest in the group (0.62) that is, they fail to detect approximately 38% of actual attacks.

Isolation Forest, on the other hand, despite having had no access whatsoever to attack labels during training, achieves the highest Recall (0.75) and the best overall F1-score (0.81), which constitutes empirical confirmation of the principal advantage of the anomaly-based approach: the ability to detect attack types that have never previously been studied.

The MLP neural network is positioned between these two extremes, achieving the second-highest AUC (0.941), which reflects the strong discriminative ability discussed above; this, however, comes at the cost of the longest training time 22 seconds as against under 1 second for the other models, which represents a clear illustration of the accuracy–adaptability–speed trade-off described in the methods-comparison table. [8]

4. Conclusion

It should be noted that, in different situations, when processing data of differing character, certain methods make it possible to resolve the problem of detecting malicious impact better than others. The signature method is effective for analyzing the content of traffic but does not permit the detection of unknown impacts. The statistical method and the artificial neural network are well suited for analyzing traffic characteristics and detecting anomalies therein, particularly for network objects characterized by stable features over time, such as, for example, servers that provide certain network services. The biometric method makes it possible to resolve the issue of information input with respect to devices according to the characteristics of their operation, and so forth. Such an impact, however, may itself constitute a discernible sign of yet another, more complex impact.

In this connection, it would be advisable to develop an approach to detecting attacks on information systems that would allow for the joint use of arbitrary methods for attack detection, taking into account their respective advantages. In many cases, however, reliance is placed not on a single specific method, but rather on the combined, joint application of the approaches enumerated above.

A specific direction for further research is the development of a hybrid Ensemble architecture combining signature-based and anomaly-based detection, in which a fast signature filter e.g., at the level of Suricata intercepts known attacks at the very first stage, while the remaining, filtered traffic is passed on to an anomaly detector based on LSTM/Autoencoder. The validity of such a model should be tested on contemporary test datasets similar to CICIDS2017/CSE-CIC-IDS2018, through the joint minimization of F1-score and FPR, with robustness verified by means of adversarial testing.

References

1. ი. ქართველიშვილი, ა. ბიჩნიგაური, ზ. ბერიძე, "აკადემიური საზოგადოების როლი კიბერუსაფრთხოების ცნობიერების ამაღლებაში," *საქართველოს ტექნიკური უნივერსიტეტის შრომები*, თბილისი: გამომცემლობა „ტექნიკური უნივერსიტეტი", 2025, ISBN 978-9941-520-86-0.
2. ი. ქართველიშვილი, ა. ბიჩნიგაური, ლ. შონია, "ფიშინგისა და მავნე კოდის მქონე ვებგვერდების პრევენციის ეფექტური მექანიზმის მოდელის შემუშავება და რეალიზაცია ვებბრაუზერის გარემოში," *საქართველოს ტექნიკური უნივერსიტეტის შრომები*, თბილისი: გამომცემლობა „ტექნიკური უნივერსიტეტი", 2023, ISBN 978-9941-512-06-3.
3. ი. ქართველიშვილი, ლ. შონია, "ვირტუალური კერძო ქსელის (VPN) აგების კონცეფცია, ქსელის ფუნქციები და მათი კლასიფიკაცია," *მართვის ავტომატიზებული სისტემები*, საქართველოს ტექნიკური უნივერსიტეტი, №2(26), თბილისი, 2018.
4. K. Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. Crown, 2014, 406 p.
5. J. Seitz, *Black Hat Python: Python Programming for Hackers and Pentesters*. No Starch Press, 2014, 171 p.
6. M. Sikorski, A. Honig, *Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software*. No Starch Press, 2012, 800 p.
7. R. Boyle, R. Panko, *Corporate Cybersecurity*. Pearson, 2025, 646 p.
8. დ. ჩაკვეტაძე, "Machine Learning მეთოდების გამოყენება ქაოსის ექსპერიმენტების შედეგების პროგნოზირებაში," ბიზნესისა და ტექნოლოგიების უნივერსიტეტი, 2026, 55 გვ.

Design and Evaluation of a Hybrid Georgian Voice Assistant for Public Services

Tamar Tutisani

Georgian American University, Tbilisi, Georgia.

Email: *tamar.tutisani@gau.edu.ge*

Nino Topuria

Candidate of Technical Sciences, Professor, Georgian Technical University, Tbilisi, Georgia.

Email: *nino.topuria@gtu.ge*

Abstract

Voice assistants have become an essential component of modern human-computer interaction, allowing users to access information through natural speech. However, most commercial voice assistants provide limited support for low-resource languages. One such language is Georgian, for which limited support for voice technologies makes it difficult for citizens to access local public services through voice interaction. This limitation is especially important for elderly users and people with visual impairments, who often use voice interfaces instead of traditional graphical user interfaces.

This paper presents the design, implementation, and evaluation of a hybrid Georgian language voice assistant that provides voice access to information about public transport and cultural events across Tbilisi. The proposed system combines cloud-based speech recognition, natural language understanding, and speech synthesis technologies within a web application architecture. Speech recognition is performed using Google Cloud Speech-to-Text Chirp 3 [3], while speech synthesis uses Microsoft Edge Neural Text-to-Speech. Intent recognition combines rule-based processing with Google Gemini 2.5 Flash, creating a hybrid architecture that improves robustness while maintaining low latency. Real-time information is obtained from TKT.ge and Google Maps services.

The developed prototype was evaluated through technical performance measurements and user-centered usability testing. Experimental results show that the proposed hybrid architecture provides reliable Georgian speech processing with acceptable latency and high usability.

Keywords: Voice Assistant, Speech-to-Text, Text-to-Speech, Intent Recognition, Natural Language Processing, Progressive Web Application, Accessibility.

Introduction

Voice-based human–computer interaction has rapidly evolved due to advances in artificial intelligence, cloud computing, and natural language processing. Modern voice assistants enable users to access digital services through natural speech, providing an alternative interaction mechanism to conventional graphical interfaces [1], [2].

Despite significant progress in speech technologies, language support remains uneven across linguistic communities. Low-resource languages often face limitations related to insufficient training data, limited availability of specialized models, and reduced integration into commercial voice platforms. Georgian represents such an environment, where the complexity of its morphological structure and limited language resources create additional challenges for speech recognition and natural language understanding systems.

The lack of effective voice-based access is particularly relevant in public information services, where users frequently need immediate access to transportation information, navigation assistance, and cultural event data. Traditional web and mobile platforms often require multiple manual interactions, which may reduce accessibility for users who prefer hands-free interaction or experience difficulties with conventional interfaces [1], [2].

Recent advances in cloud-based speech technologies and large language models have improved the capabilities of conversational systems [3]–[5]. However, relying exclusively on large language models for public information services may increase latency and computational costs.

Therefore, efficient voice assistants require hybrid approaches that combine deterministic processing for common tasks with advanced language understanding for complex interactions. This research presents the design and evaluation of a hybrid Georgian-language voice assistant developed for public transportation and cultural information services in Tbilisi. The proposed system investigates how existing cloud AI technologies can be adapted to support natural speech interaction in a low-resource language environment.

The main contributions of this work are summarized as follows:

1. A hybrid cloud-based architecture for Georgian-language voice interaction integrating speech recognition, intent processing, information retrieval, and speech synthesis.
2. A hybrid intent recognition approach combining rule-based processing with Google Gemini 2.5 Flash to improve efficiency and contextual understanding.
3. Georgian-specific text normalization techniques for improving speech synthesis quality.
4. Integration of real-time public information sources, including transportation and cultural event data.
5. Experimental evaluation based on technical performance measurements and usability assessment using the System Usability Scale (SUS) [6].

System Architecture and Implementation

Within this research, a Georgian-language voice assistant prototype, named the “Tbilisi Assistant”, was developed to provide users with information about public transportation and cultural events through natural speech interaction. The system aims to improve access to public information by reducing reliance on conventional text-based interfaces. It adopts a three-layer architecture comprising the user interface layer, application logic layer, and data management layer, which enhances modularity, supports independent component development, and facilitates future system expansion.

The user interface is implemented as a Progressive Web Application (PWA), providing cross-platform accessibility without requiring separate installation. The interface follows a mobile-first design approach [7] and incorporates accessibility-oriented features, including high-contrast visual elements, screen-reader compatibility, and voice-centered interaction scenarios [8],[9].

The backend layer is implemented using the Python FastAPI framework, which provides asynchronous request processing and efficient communication with external services. The backend manages audio processing, user request analysis, information retrieval, and response generation. SQLite is used for temporary storage of cached transportation and cultural information.

The voice processing pipeline begins with speech recognition, where user audio input is captured through browser-based media interfaces [10] and converted into text. During system development, local Whisper-based processing was evaluated; however, the obtained Georgian recognition quality and CPU execution time were insufficient for real-time interaction requirements. Therefore, the final implementation uses Google Cloud Speech-to-Text Chirp 3 [3], which provides more stable recognition performance for Georgian speech, including local names and geographic terms.

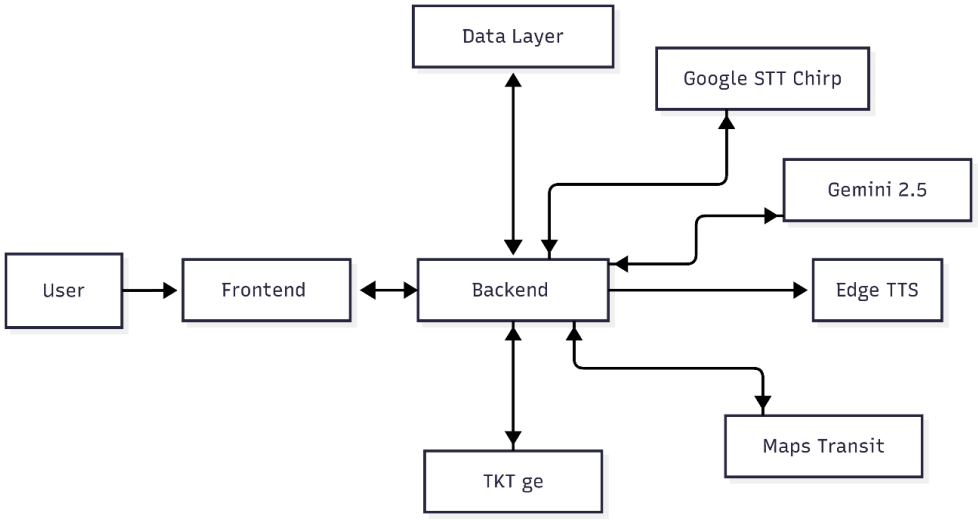
User intent recognition is implemented through a hybrid NLP approach combining rule-based processing and Google Gemini 2.5 Flash [4]. Frequently occurring structured queries, such as finding nearby stops or requesting route information, are processed using predefined rules to reduce latency and minimize unnecessary large language model requests. More complex natural language queries are forwarded to Gemini for contextual interpretation.

To address Georgian linguistic characteristics, preprocessing and text normalization techniques were applied. These techniques handle morphological variations, date and time expressions, abbreviations, and grammatical structures, improving both intent interpretation and speech output quality. Speech responses are generated using Microsoft Edge Neural Text-to-

Speech [5]. Before synthesis, generated responses undergo normalization to ensure more natural Georgian pronunciation.

The system integrates external data sources for real-time information retrieval. Transportation-related queries use Google Maps Platform services, including Directions API, Places API, and Geocoding API [11]. Cultural event information is obtained from the TKT.ge platform. To optimize response speed and data freshness, a stale-while-revalidate caching strategy is applied, allowing immediate access to cached information while updating data in the background. The complete processing pipeline is illustrated in Figure 1. The workflow includes audio acquisition, speech recognition, intent classification, information retrieval, response generation, and speech synthesis.

Figure 1. System architecture of the "Tbilisi Assistant"



Research Methodology and Evaluation Approach

The research was conducted using a system design and prototype-based development methodology. The objective was to design and evaluate a Georgian-language voice interface capable of providing public information services through natural speech interaction.

Initially, functional and non-functional requirements were defined. Functional requirements included voice input processing, speech recognition, intent identification, information retrieval, and voice response generation. The selected application scenarios focused on public transportation and cultural event information. Non-functional requirements addressed practical deployment considerations, including response latency, usability, accessibility, platform independence, and operational efficiency.

The development process consisted of several stages. First, different technological components were evaluated, including local and cloud-based speech recognition approaches. Second, the prototype architecture was implemented and integrated with external services. Finally, technical performance evaluation and user-centered usability testing were conducted.The technical evaluation focused on response latency, speech recognition stability, and overall system performance. User experience was assessed using the System Usability Scale (SUS), providing a quantitative measure of interface usability [6].

Architectural Design Rationale

The architectural design was driven by the need to provide an efficient and accessible voice interaction framework while maintaining flexibility for future extensions. The PWA-based interface was chosen to ensure platform independence and simplify deployment across different devices without requiring separate native applications. This approach also supports the integration of accessibility features required for inclusive public service applications.

The hybrid intent recognition mechanism was introduced to address the trade-off between response speed and natural language understanding. Structured requests, such as transportation searches and route-related queries, are processed using predefined rules to achieve low latency, while more complex user inputs are analyzed using the large language model component to improve contextual interpretation.

Cloud-based AI services were selected to provide reliable speech recognition and synthesis capabilities without requiring significant local computational resources. This approach is particularly appropriate for Georgian, where developing and maintaining fully local speech models remains challenging due to limited language resources.

The modular architecture enables future integration of additional public services, improved Georgian language models, and new data sources without significant changes to the existing system structure.

Conclusion

This study presented the design, implementation, and evaluation of a hybrid Georgian-language voice assistant for public transportation and cultural information services. The research demonstrated the feasibility of adapting existing artificial intelligence technologies to support natural speech interaction in a low-resource language context.

The experimental evaluation showed that the combination of cloud-based speech services, hybrid intent recognition, and Georgian-specific language processing techniques provides an effective solution for practical voice-based public information systems. The proposed approach enabled efficient processing of common requests while preserving the ability to interpret more complex natural language queries. The system achieved a SUS score of 89.17, indicating strong perceived usability among evaluation participants.

The developed prototype demonstrates that combining modern AI services with accessibility-oriented design principles can improve access to digital public services for Georgian-speaking users. The modular architecture also provides a foundation for future extensions, including additional service domains, enhanced language models, and multilingual support.

The study has several limitations. The current implementation relies on external cloud services and stable Internet connectivity, and the user evaluation was conducted with a limited

participant group. Future research will focus on expanding real-world testing, improving Georgian language understanding capabilities, and investigating more adaptive conversational models for public service applications.

REFERENCES

[1] M. T. Harris and W. A. Rogers, "Social participation challenges and response strategies identified by adults aging with a disability", *Journal of Elder Policy*, vol. 3, pp. 95–109, 2024.

[2] P. Esquivel, K. Gill, M. Goldberg, S. A. Sundaram, L. Morris, and D. Ding, "Voice Assistant Utilization among the Disability Community for Independent Living: A Rapid Review of Recent Evidence, "*Human Behavior and Emerging Technologies*, Article no. 6494944, 2024.

[3] Google Cloud, "Speech-to-Text documentation", *Google Cloud Documentation*. https://cloud.google.com/speech-to-text/docs

[4] Google, "Gemini API documentation", *Google AI for Developers*. https://ai.google.dev/gemini-api/docs

[5] Microsoft, "Speech service documentation", *Microsoft Azure Documentation*. https://learn.microsoft.com/en-us/azure/ai-services/speech-service/

[6] J. Sauro and J. R. Lewis, Quantifying the User Experience: Practical Statistics for User Research, 3rd ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[7] A. Firtman, Progressive Web Apps: The Future of Web Development, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2023.

[8] W3C, "Web Content Accessibility Guidelines (WCAG) 2.1", World Wide Web Consortium. https://www.w3.org/TR/WCAG21/

[9] W3C, "Accessible Rich Internet Applications (WAI-ARIA) 1.2," World Wide Web Consortium. https://www.w3.org/TR/wai-aria-1.2/

[10] Mozilla Developer Network, "MediaDevices.getUserMedia", *MDN Web Docs*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/MediaDevices/getUserMedia

[11] Microsoft, "What is a Progressive Web App?", *Microsoft Edge Documentation*. https://learn.microsoft.com/en-us/microsoft-edge/progressive-web-apps-chromium/

[12] Google, "Google Maps Platform Documentation," *Google for Developers*. https://developers.google.com/maps/documentation

AI-Assisted Editorial Decision Making in Open Journal Systems (OJS): A Framework for Manuscript Scope Analysis and Reviewer Recommendation

Besiki Tabatadze

PhD (Applied Mathematics), Affiliated Professor, European University, Tbilisi, Georgia.

Email: *tabatadze.besik@eu.edu.ge.*

Teimuraz Sturua

PhD (Mathematician), Professor, European University, Tbilisi, Georgia.

Email: *sturua.teimuraz@eu.edu.ge.*

Nika Tabatadze

MSc (Informatics), Invited Lecturer, European University, Tbilisi, Georgia.

Email*: nikatabatadze9@gmail.com.*

Ekaterine Natsvlishvili

PhD (Social Philosophy), MA (Management), Affiliated Professor, European University, Tbilisi, Georgia.

Email: *natsvlishvili.ekaterine@eu.edu.ge*

Abstract

The increasing volume of manuscript submissions has created significant challenges for editorial boards in maintaining efficient and consistent editorial workflows. Initial manuscript screening, assessment of manuscript relevance to the journal scope, and identification of appropriate reviewers are among the most time-consuming tasks in academic publishing. Although Open Journal Systems (OJS) provides a comprehensive platform for managing editorial processes, these tasks are still performed primarily through manual evaluation.

This study presents an AI-assisted editorial support framework designed for Open Journal Systems (OJS). The proposed approach applies natural language processing techniques to analyze manuscript titles, abstracts, keywords, and reviewer expertise profiles. Textual information is transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) vectorization method, while cosine similarity is employed to measure

semantic relevance between submitted manuscripts, journal scope descriptions, and reviewer expertise. Based on these similarity scores, the system recommends the most relevant journal scope categories and suitable reviewers for each submitted manuscript.

The proposed framework is implemented as a Python-based prototype and demonstrates the feasibility of integrating artificial intelligence into editorial workflows without replacing the role of the editor. Instead, the system functions as an intelligent decision-support tool that provides transparent and consistent recommendations during the initial manuscript screening process. The proposed approach has the potential to reduce editorial workload, improve reviewer assignment, and enhance the overall efficiency of manuscript management within Open Journal Systems.

Keywords: Artificial Intelligence, Open Journal Systems (OJS), Natural Language Processing, TF-IDF, Cosine Similarity, Manuscript Scope Analysis, Reviewer Recommendation.

Introduction

Artificial intelligence (AI) has become one of the most influential technologies in contemporary science and engineering, significantly transforming numerous domains, including healthcare, education, finance, manufacturing, and scientific publishing. Recent advances in natural language processing (NLP), information retrieval, and intelligent text analysis have created new opportunities for automating knowledge-intensive tasks that traditionally required substantial human expertise. These developments have also influenced academic publishing, where AI technologies are increasingly being investigated as tools for supporting editorial workflows, improving manuscript processing, and enhancing the overall efficiency of scholarly communication (LeCun et al., 2015; OpenAI, 2023).

The continuous growth in scientific research has led to a rapid increase in manuscript submissions received by academic journals. As a consequence, editorial boards face growing workloads associated with the initial screening of submitted manuscripts, evaluation of their

relevance to the journal's aims and scope, identification of qualified reviewers, and coordination of the peer-review process. Although editorial management platforms have significantly improved the organization of publishing activities, many critical decisions still rely on manual assessment by editors. This process is often time-consuming, may introduce inconsistencies, and becomes increasingly difficult as submission volumes continue to grow (Smith, 2006; Minciună & Țurcan, 2025).

Among the available editorial management platforms, Open Journal Systems (OJS) has become one of the most widely adopted open-source solutions for managing scholarly journals. Developed by the Public Knowledge Project (PKP), OJS provides an integrated environment supporting manuscript submission, editorial workflow management, peer review, copyediting, production, and publication. Its flexibility, accessibility, and extensive international adoption have made it a preferred platform for universities, research institutions, and independent publishers. Nevertheless, despite its comprehensive workflow management capabilities, OJS currently provides limited intelligent support for evaluating manuscript relevance to journal scope or recommending appropriate reviewers during the initial editorial screening process (Willinsky, 2005; Edgar & Willinsky, 2010; Tabatadze, 2024).

The initial editorial assessment represents one of the most important stages of scholarly publishing. Before a manuscript enters the peer-review process, editors must determine whether its topic corresponds to the journal's thematic scope and whether the manuscript should proceed to external review. Subsequently, appropriate reviewers must be selected based on their research expertise and scientific interests. These activities require careful analysis of manuscript content and reviewer profiles and are largely dependent on editorial experience. Consequently, the development of intelligent support systems capable of analyzing textual similarity between manuscripts, journal scope descriptions, and reviewer expertise has become an important research direction in academic publishing.

Recent advances in natural language processing and information retrieval provide effective methods for addressing these challenges. Techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) enable textual documents to be transformed into numerical vector representations, while cosine similarity offers an interpretable measure for estimating semantic similarity between textual documents. These approaches have been successfully applied in numerous text classification and document retrieval tasks and provide a transparent foundation for supporting editorial recommendations without replacing human decision-making (Manning et al., 2008; Jurafsky & Martin, 2009; Sebastiani, 2002; Kowsari et al., 2019).

This study proposes an AI-assisted editorial support framework for Open Journal Systems (OJS) that focuses on two essential tasks of the initial editorial workflow: manuscript scope analysis and reviewer recommendation. The proposed framework applies TF-IDF vectorization and cosine similarity to analyze manuscript titles, abstracts, keywords, journal scope descriptions, and reviewer expertise profiles. Rather than replacing editorial judgment, the proposed system functions as an intelligent decision-support tool that provides transparent recommendations during manuscript screening and reviewer selection. The remainder of this paper is organized as follows. Section 2 reviews the editorial workflow in Open Journal Systems and discusses its current challenges. Section 3 presents the proposed AI-assisted framework and describes its architecture and implementation. Section 4 demonstrates the experimental results obtained using the developed prototype, while the final section summarizes the main findings and outlines directions for future research.

1. Editorial Workflow in Open Journal Systems

1.1. Traditional Editorial Workflow

The editorial workflow of an academic journal consists of a sequence of interconnected stages designed to ensure the quality, integrity, and transparency of scholarly publishing. Modern

journal management systems provide structured environments that facilitate communication among authors, editors, reviewers, and publishers throughout the publication process. Among these systems, Open Journal Systems (OJS) has become one of the most widely adopted open-source platforms for managing scholarly journals due to its flexibility, scalability, and comprehensive support for editorial activities (Willinsky, 2005; Edgar & Willinsky, 2010).

The editorial process begins with manuscript submission, during which authors upload their manuscripts together with essential metadata, including the title, abstract, keywords, author information, and subject classification. These metadata provide the initial descriptive information required for editorial management and subsequent manuscript evaluation. Once the submission is completed, the manuscript enters the editorial workflow and becomes available for assessment by the editorial office.

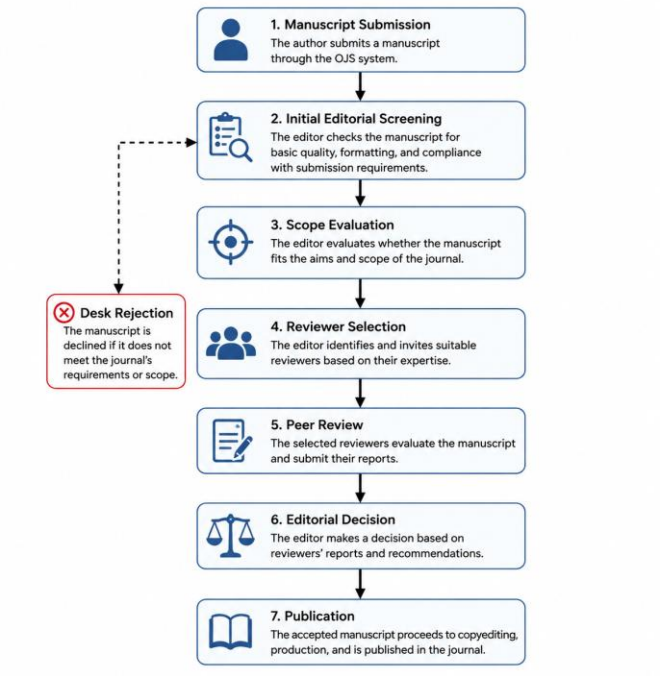
The first stage of editorial evaluation is the initial screening, where editors verify whether the submitted manuscript satisfies the journal's formal requirements and falls within the journal's thematic scope. During this stage, editors assess manuscript completeness, compliance with submission guidelines, and relevance to the journal's research areas. Manuscripts that do not correspond to the aims and scope of the journal are typically rejected before entering the peer-review process, thereby reducing unnecessary workload for reviewers and improving the overall efficiency of editorial management (Smith, 2006; Tabatadze, 2024).

Manuscripts that successfully pass the initial screening proceed to the reviewer selection stage. Selecting appropriate reviewers is one of the most important responsibilities of editors because the quality, fairness, and scientific reliability of peer review largely depend on reviewer expertise. Editors are therefore required to identify reviewers whose research interests and scientific experience closely correspond to the subject of the submitted manuscript. This process commonly involves manual examination of reviewer profiles, publication records, and areas of expertise, making reviewer assignment both time-consuming and dependent on editorial experience (Minciună & Țurcan, 2025).

After reviewers are assigned, the manuscript enters the peer-review stage, where independent experts evaluate its originality, methodological quality, scientific contribution, and overall suitability for publication. Reviewer reports and recommendations support the editor in making the final editorial decision, which may result in manuscript acceptance, requests for revision, or rejection. Throughout the entire editorial workflow, responsibility for the final decision remains exclusively with the editor.

Although Open Journal Systems effectively supports the administrative management of editorial workflows, the platform currently provides limited intelligent assistance during the initial manuscript screening and reviewer selection stages. These activities continue to rely primarily on manual analysis of manuscript content and reviewer expertise. As submission volumes continue to increase across academic journals, there is a growing need for intelligent support systems capable of assisting editors in these early stages of the editorial workflow. This need provides the foundation for the AI-assisted framework proposed in this study (Tabatadze, 2024).

Figure 1. *Traditional Editorial Workflow in Open Journal Systems*



1.2. Challenges of Traditional Editorial Workflow

The continuous expansion of scientific research has led to a substantial increase in the number of manuscripts submitted to academic journals worldwide. While this growth reflects the dynamic development of global research communities, it also places considerable pressure on editorial boards responsible for managing increasingly complex publication workflows. Editors are expected to process submissions efficiently while maintaining high standards of scientific quality, transparency, fairness, and consistency throughout the editorial process. Consequently, the growing volume of manuscript submissions has become one of the major challenges facing contemporary scholarly publishing (Smith, 2006; Minciună & Țurcan, 2025).

One of the primary challenges in the editorial workflow is the initial assessment of a manuscript's relevance to the journal's aims and scope. Before initiating the peer-review process, editors must determine whether the research topic, objectives, methodology, and expected contribution correspond to the thematic areas covered by the journal. Manuscripts that clearly fall outside the journal's scope are generally rejected during the initial editorial screening, thereby reducing unnecessary reviewer workload and improving the efficiency of the editorial process. However, assessing thematic relevance requires careful examination of manuscript titles, abstracts, keywords, and subject classifications, making this stage both time-consuming and highly dependent on editorial expertise (Willinsky, 2005; Tabatadze, 2024b).

Another significant challenge is the identification of appropriate reviewers. The effectiveness of peer review depends largely on assigning manuscripts to experts whose scientific background and research interests closely correspond to the submitted work. In practice, editors often examine reviewer databases, publication records, institutional profiles, and declared areas of expertise to identify suitable reviewers. As reviewer databases continue to expand and scientific disciplines become increasingly specialized and interdisciplinary, selecting the most appropriate reviewers becomes progressively more difficult. Furthermore, reviewer availability, previous review assignments, workload, and potential conflicts of interest introduce additional

complexity into the reviewer selection process (Edgar & Willinsky, 2010; Minciună & Țurcan, 2025).

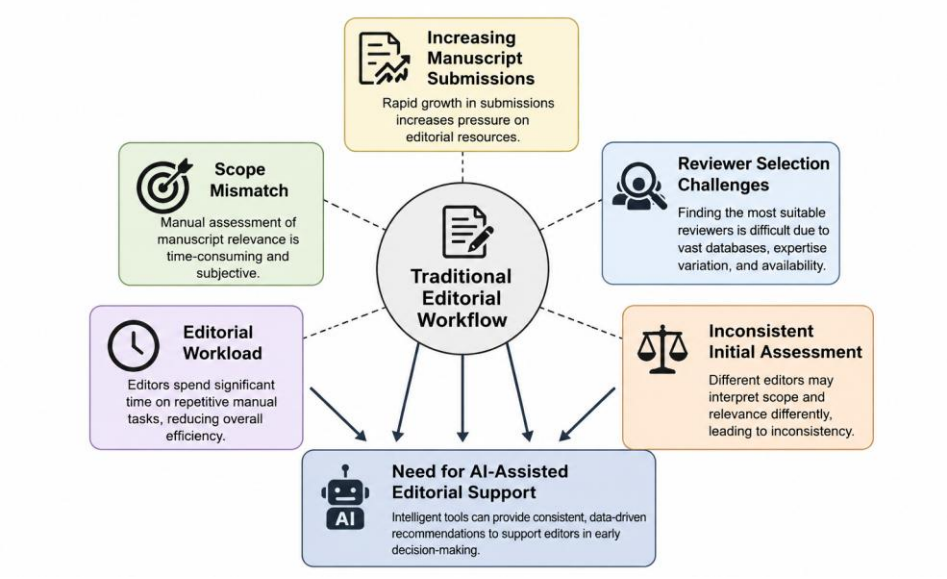
Maintaining consistency during the initial editorial assessment represents another important challenge. Different editors may evaluate manuscript relevance, reviewer suitability, and journal scope differently depending on their disciplinary knowledge, professional experience, and interpretation of editorial policies. Although editorial guidelines establish general evaluation criteria, professional judgment remains an essential component of the editorial process. Consequently, manuscripts with similar characteristics may receive different initial assessments, highlighting the need for more transparent and consistently applied decision-support mechanisms.

Recent advances in natural language processing and information retrieval provide practical opportunities for supporting editors during the early stages of manuscript evaluation. Text-based analytical methods enable the comparison of manuscript content with journal scope descriptions and reviewer expertise profiles, providing valuable evidence for editorial decision support. Techniques such as TF-IDF vectorization and cosine similarity have demonstrated their effectiveness in document representation, text comparison, and information retrieval applications, making them well suited for supporting manuscript scope evaluation and reviewer recommendation (Manning et al., 2008; Jurafsky & Martin, 2009; Sebastiani, 2002; Kowsari et al., 2019).

The challenges discussed above demonstrate the growing need for intelligent editorial support tools capable of assisting editors during the initial stages of manuscript processing. Rather than replacing editorial judgment, such systems are intended to provide transparent, data-informed recommendations that support scope evaluation and reviewer selection while preserving the editor's responsibility for all editorial decisions. Motivated by these challenges, the following section presents the proposed AI-assisted editorial framework for Open Journal

Systems (OJS), which applies natural language processing techniques together with TF-IDF vectorization and cosine similarity analysis to support the initial editorial workflow.

Figure 2. *Challenges of Traditional Editorial Workflow*



2. Proposed AI-Assisted Editorial Framework

2.1. Overall Framework

The AI-assisted editorial framework proposed in this study is designed to support editors during the initial stages of the editorial workflow in Open Journal Systems (OJS). Unlike conventional editorial management systems, which primarily facilitate workflow administration, the proposed framework integrates manuscript scope analysis and reviewer recommendation into a unified AI-assisted decision-support process. Rather than replacing human judgment, the framework provides intelligent recommendations that assist editors in evaluating manuscript relevance to the journal's scope and identifying reviewers whose expertise closely matches the submitted research. Its primary objective is to improve the efficiency, consistency, and

transparency of the editorial process while preserving the editor's full responsibility for all editorial decisions.

The proposed framework consists of several interconnected components responsible for processing information extracted from submitted manuscripts, journal scope descriptions, and reviewer expertise profiles. The primary input includes manuscript titles, abstracts, keywords, journal scope descriptions, and reviewer research interests. Before similarity analysis, the collected textual information undergoes a sequence of natural language processing procedures, including preprocessing and feature extraction, and is subsequently transformed into numerical vector representations suitable for computational analysis.

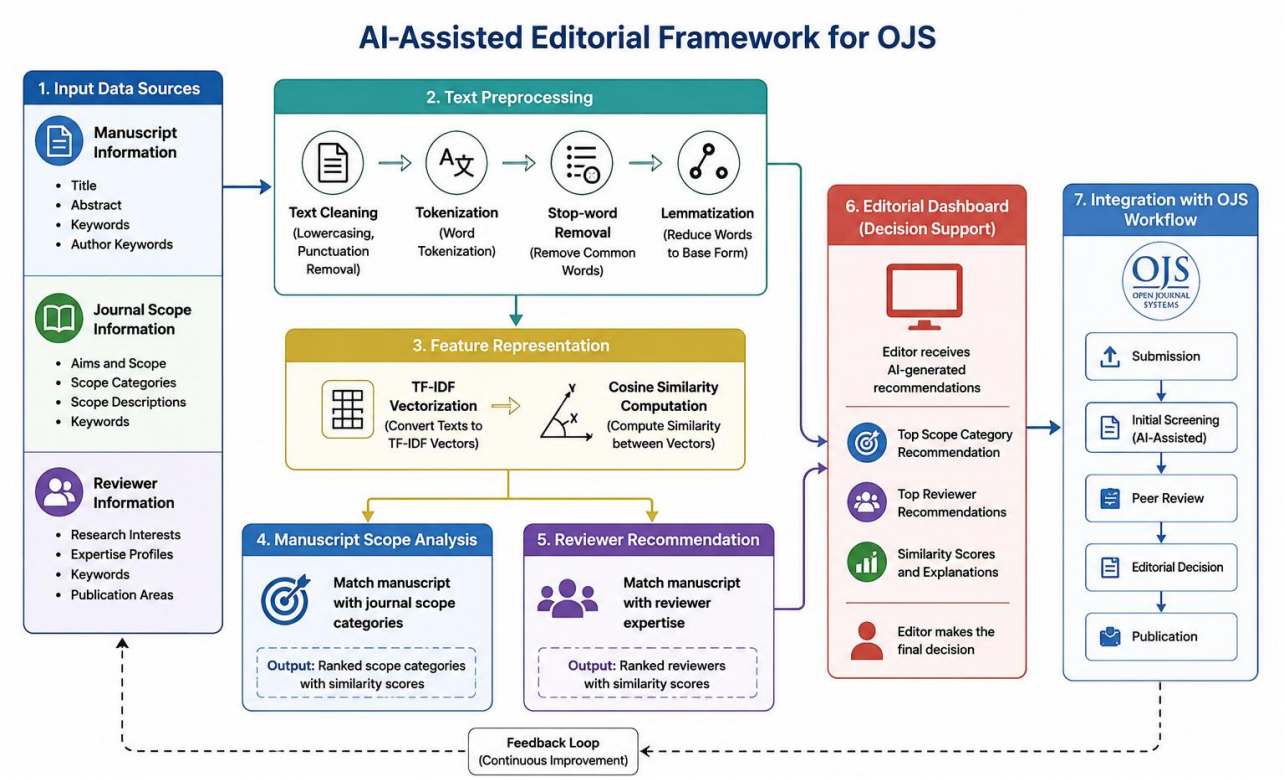
To represent textual documents numerically, the framework applies the Term Frequency–Inverse Document Frequency (TF-IDF) vectorization technique. Semantic similarity between manuscripts, journal scope descriptions, and reviewer expertise profiles is then calculated using cosine similarity. The resulting similarity scores indicate how closely a submitted manuscript corresponds to the journal's thematic areas and how well individual reviewers match the manuscript based on their research expertise.

Based on the calculated similarity scores, the framework produces two complementary outputs. The first identifies the journal scope category that best corresponds to the submitted manuscript, supporting editors during the initial screening stage. The second ranks potential reviewers according to the similarity between manuscript content and reviewer expertise profiles. These outputs provide interpretable decision-support information that assists editors in making consistent and informed decisions while preserving full editorial control over manuscript evaluation and reviewer assignment.

The proposed framework has been implemented as a Python-based prototype using widely adopted natural language processing and machine learning libraries. The implementation demonstrates that AI-based decision-support techniques can be integrated into existing Open Journal Systems without modifying the core architecture of the platform. Consequently, the

framework can function as an auxiliary intelligent module that complements the existing editorial workflow and enhances the efficiency of manuscript processing.

Figure 3. *Architecture of the Proposed AI-Assisted Editorial Framework.*



2.2. Dataset and Text Processing

The proposed AI-assisted framework operates on three categories of textual data: submitted manuscripts, journal scope descriptions, and reviewer expertise profiles. These datasets provide the information required to support manuscript scope analysis and reviewer recommendation during the initial stages of the editorial workflow.

To evaluate the proposed framework, an editorial dataset was constructed combining actual and simulated records. The dataset consists of 100 manuscript records, of which 17 correspond to actual submitted manuscripts and the remaining 83 are simulated; 11 journal scope categories, all based on the actual thematic areas covered by the journal; and 18 reviewer profiles, of which 7 correspond to actual reviewers and the remaining 11 are simulated. The manuscript

dataset includes titles, abstracts, and author-provided keywords representing submitted research papers. The journal scope dataset contains thematic descriptions defining the research areas covered by the journal, while reviewer profiles describe research interests, scientific expertise, and representative keywords used for reviewer matching.

The real editorial data used in this study were obtained from the editorial database of the Computational and Applied Science Journal (CAS Journal). All real records were anonymized prior to analysis, and additional simulated records were incorporated to create a combined experimental dataset.

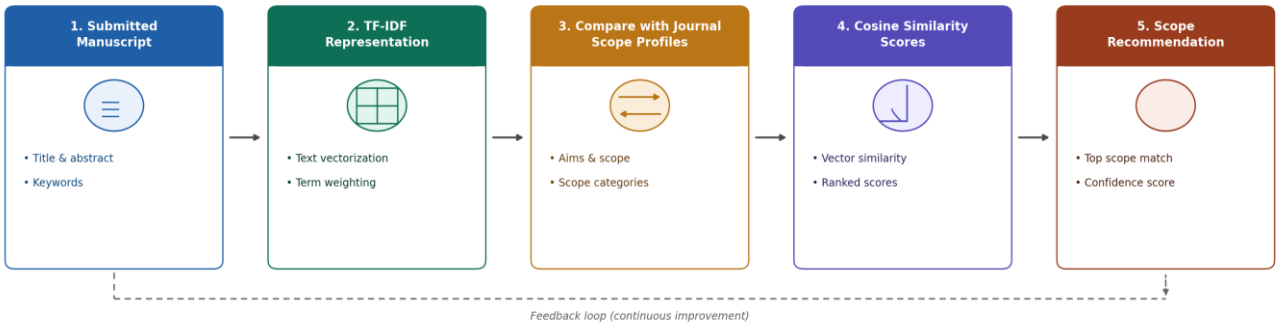
Before similarity analysis, all textual data undergo a preprocessing stage designed to improve the consistency and quality of document representation. The preprocessing pipeline includes text normalization, conversion to lowercase, tokenization, removal of punctuation and common stop words, and whitespace normalization. These procedures reduce textual noise and prepare the documents for numerical feature extraction.

Following preprocessing, the cleaned textual information is transformed into numerical representations that serve as the basis for manuscript scope evaluation and reviewer recommendation. The characteristics of the experimental dataset are summarized in Table 1, while the preprocessing workflow is presented in Figure 4.

Table 1. *Summary of the Experimental Dataset*

Dataset	Number of Records	Description
Manuscripts	100	17 actual and 83 simulated manuscript titles, abstracts, and keywords
Journal Scope Categories	11	Actual research areas and scope descriptions
Reviewer Profiles	18	7 actual and 11 simulated reviewer expertise profiles and research interests

Figure 4. *Text Processing Pipeline*



2.3. Manuscript Scope Analysis

The primary objective of manuscript scope analysis is to determine the degree of correspondence between a submitted manuscript and the thematic areas covered by the journal. This stage supports editors during the initial screening process by providing an objective similarity measure based on the textual content of the manuscript and the journal's scope descriptions.

Following text preprocessing, all manuscript abstracts, journal scope descriptions, and keywords are transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) representation. TF-IDF assigns weights to terms according to their importance within a document while reducing the influence of frequently occurring words that contribute little semantic information (Manning et al., 2008).

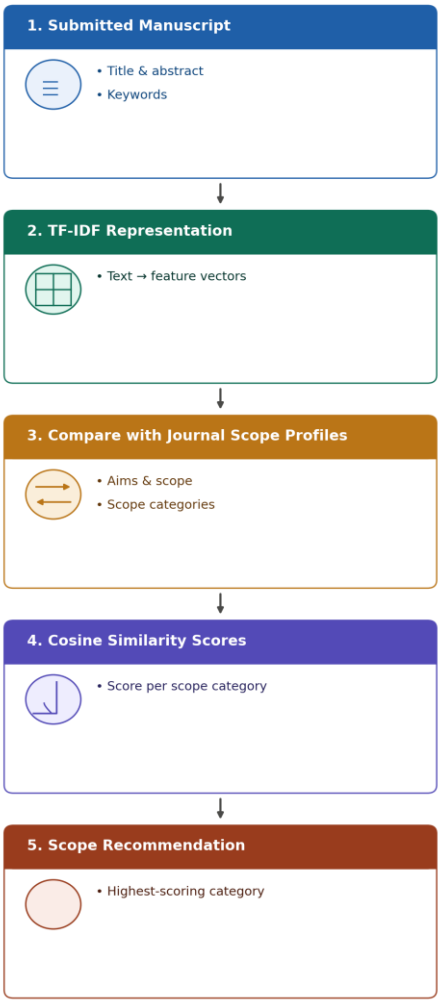
After vectorization, the similarity between the submitted manuscript and each journal scope category is calculated using cosine similarity. This metric evaluates the angular similarity between two feature vectors and produces a similarity score ranging from 0 to 1, where larger values indicate stronger semantic correspondence (Jurafsky & Martin, 2009; Sebastiani, 2002).

For each submitted manuscript, similarity scores are calculated against all journal scope categories. The category with the highest similarity score is identified as the most appropriate

thematic area for the manuscript. These results provide editors with quantitative evidence supporting the initial scope evaluation while preserving editorial discretion in the final decision.

The manuscript scope analysis process implemented in the proposed framework is illustrated in Figure 5.

Figure 5. *Manuscript Scope Analysis Process*



2.4. Reviewer Recommendation

The reviewer recommendation module is designed to assist editors in identifying reviewers whose expertise most closely corresponds to the subject and content of a submitted

manuscript. Appropriate reviewer selection is essential for maintaining the quality, relevance, and reliability of the peer-review process.

Within the proposed framework, each reviewer is represented by a textual profile containing primary and secondary areas of expertise, research interests, and expertise-related keywords. Reviewer profiles undergo the same preprocessing and vectorization procedures applied to manuscript data, ensuring a consistent numerical representation of both manuscripts and reviewer expertise.

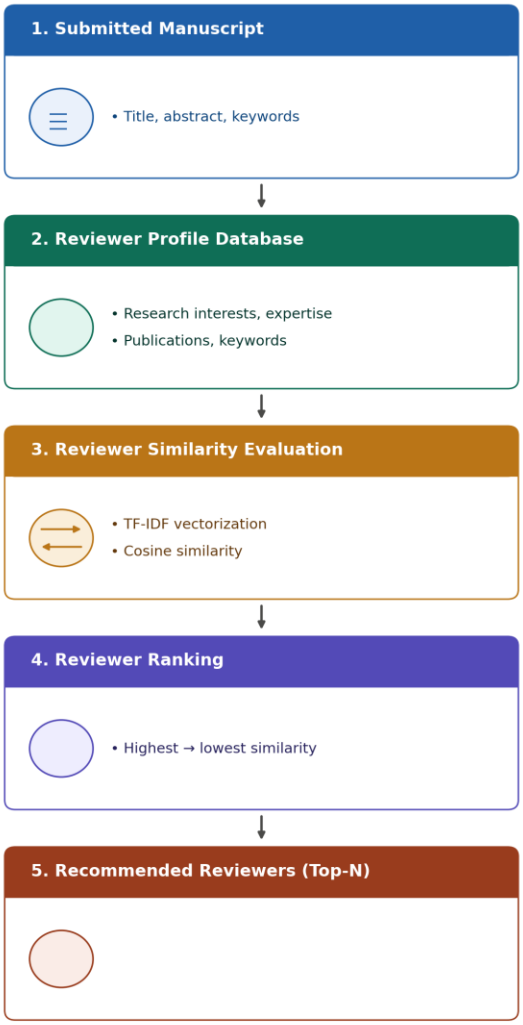
After TF-IDF vectorization, cosine similarity is calculated between the manuscript vector and each reviewer profile vector. The resulting similarity scores indicate the degree of correspondence between the manuscript content and the reviewer's expertise. Reviewers are then ranked in descending order according to these scores, enabling the system to recommend the most relevant candidates for the submitted manuscript.

The recommendation process is based primarily on thematic similarity and is intended to support, rather than replace, editorial judgment. Before assigning a reviewer, editors must also consider additional factors that are not captured by textual similarity alone, including reviewer availability, workload, previous assignments, institutional affiliation, and potential conflicts of interest. Therefore, the generated ranking should be interpreted as decision-support information rather than as an automatic reviewer assignment.

By narrowing the initial pool of potential reviewers, the proposed module can reduce the time required for reviewer identification and improve the consistency of reviewer selection. Final responsibility for reviewer invitation and assignment remains entirely with the editor.

The reviewer recommendation process is illustrated in Figure 6.

Figure 6. *Reviewer Recommendation Process*



2.5. Framework Demonstration

To demonstrate the practical applicability of the proposed framework, a Python-based prototype was developed that integrates manuscript scope analysis and reviewer recommendation into a single decision-support workflow. The prototype accepts manuscript metadata, including the title, abstract, and keywords, and processes the submitted text using the methodology described in the previous sections.

During execution, the system first determines the journal scope category that best corresponds to the submitted manuscript by calculating similarity scores between the manuscript

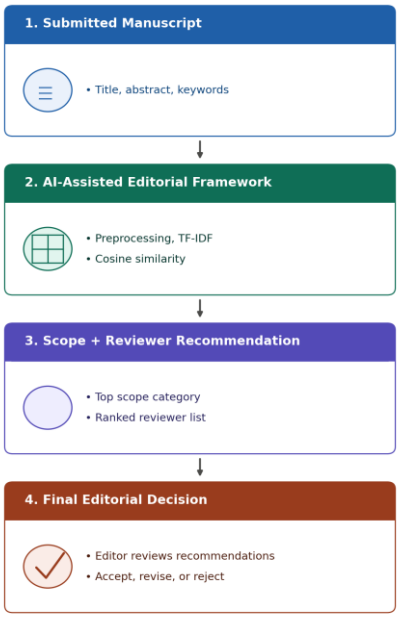
and predefined journal research areas. Subsequently, the same manuscript is compared with reviewer expertise profiles, producing a ranked list of potential reviewers according to thematic similarity.

The generated recommendations are intended to assist editors during the initial editorial assessment by providing objective and interpretable similarity scores. Rather than replacing editorial expertise, the framework supports decision-making by reducing the time required for manuscript screening and reviewer identification while improving the consistency of the editorial workflow.

The prototype demonstrates that artificial intelligence techniques based on natural language processing and information retrieval can be effectively integrated into Open Journal Systems as an auxiliary decision-support component without modifying the existing editorial process.

An example of the proposed framework and its generated recommendations is presented in Figure 7.

Figure 7. *Framework Demonstration*



Conclusion

The increasing volume of manuscript submissions has intensified the workload of editorial boards and created a growing need for reliable tools that can support the initial stages of manuscript evaluation. This study proposed an AI-assisted editorial framework for Open Journal Systems (OJS) that combines manuscript scope analysis and reviewer recommendation within a unified decision-support process.

The framework applies natural language processing techniques, TF-IDF vectorization, and cosine similarity to compare manuscript content with journal scope descriptions and reviewer expertise profiles. As a result, it generates similarity-based recommendations that can assist editors in assessing the thematic relevance of submitted manuscripts and identifying reviewers whose expertise most closely corresponds to the research topic.

The main contribution of the study is the integration of these two editorial tasks into a single, transparent framework designed to complement the existing OJS workflow. The proposed approach does not replace editorial judgment or automate final decisions. Instead, it provides interpretable support during manuscript screening and reviewer selection, while responsibility for all editorial decisions remains with the editor.

The Python-based prototype and the combined dataset of actual and simulated editorial records demonstrate the feasibility of the proposed approach. However, the current framework is limited to text-based similarity analysis and does not automatically consider reviewer availability, workload, institutional relationships, previous performance, or conflicts of interest. These factors must continue to be evaluated by the editorial team.

Future research will focus on testing the framework using real editorial data, expanding reviewer-profile information, and comparing TF-IDF-based similarity with more advanced semantic representation methods. Further development may also include implementation of the framework as an OJS-compatible module and evaluation of its effect on processing time, recommendation quality, and editorial consistency.

Overall, the proposed framework provides a practical foundation for introducing AI-assisted decision support into Open Journal Systems. It demonstrates how transparent text-analysis methods can contribute to more efficient and consistent editorial workflows while maintaining human oversight throughout the scholarly publishing process.

REFERENCES

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.

Edgar, B., & Willinsky, J. (2010). *A survey of the scholarly journals using Open Journal Systems*. *Scholarly and Research Communication, 1*(2), 1–22.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing* (2nd ed.). Pearson.

Kowsari, K., Heidarysafa, M., Brown, D. E., Meimandi, K. J., & Barnes, L. E. (2019). Text classification algorithms: A survey. *Information, 10*(4), 150. https://doi.org/10.3390/info10040150

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*. Dartmouth College.

Minciună, V., & Țurcan, N. (2025). Journal editors' views on scientific publishing in the Republic of Moldova: Results of a national survey. *Revista Română de Biblioteconomie și Știința Informării, 21*(1), 22–40.

OpenAI. (2023). *GPT-4* technical *report*. *arXiv*. https://doi.org/10.48550/arXiv.2303.08774

Owen, B. (2012). The *Public Knowledge Project and Open Journal Systems*. *PKP Scholarly Publishing Conference Proceedings*.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0

Safitri, R. M. (2026). Analysis of risk factors for computer vision syndrome (CVS) among journal management staff utilizing Open Journal Systems (OJS). *Smart: Journal of Healthcare, 1*(1), 68–92.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1–47. https://doi.org/10.1145/505282.505283

Smith, R. (2006). The trouble with medical journals. *Journal of the Royal Society of Medicine, 99*(3), 115–119.

Tabatadze, B. (2024). Prospects of using AI technologies in Open Journal Systems (OJS). *Journal of Technical Science and Technologies, 8*(2), 68–74.

Tabatadze, B. (2024). Technological aspects of Open Journal Systems (OJS). *Journal of Technical Science and Technologies, 8*(1), 23–29.

Tabatadze, B. (2026). AI-assisted programming in practice: Conceptual foundations and data-informed observations. *Computational and Applied Science, 1*(1), 84–100.

Tabatadze, B., & Sokhadze, E. (2023). Possibility of improving the training courses content based on statistical analysis of the collected points. *Globalization and Business, 8*(15), 54–56.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Willinsky, J. (2005). Open Journal Systems: An example of open source software for journal management and publishing. *Library Hi Tech, 23*(4), 504–519.

An AI-Driven Multimethod Framework for Inclusive Education: A Cross-Regional Survey Analysis

Mariam Chkhaidze

PhD (Technical Science), PhD (Educational Science), Professor, Geoorgian Technical University, Tbilisi, Georgia.

Email: *m.chkhaidze@gtu.ge.*

Levan Mateshvili

PhD (Educational Science), PhD (History), PhD (Theology), Professor, Geoorgian Technical University, Tbilisi, Georgia.

Email: *l.mateshvili74@gmail.com.*

Abstract

The integration of Artificial Intelligence (AI) into educational science offers transformative opportunities to advance inclusive education, particularly for students with visual impairments. This study presents an AI-driven analysis of surveys conducted across European and Georgian higher education institutions, drawing on responses from students, faculty, medical professionals, and psychologists. Using a mixed-methods design, the research combines traditional statistical analysis with advanced AI methodologies, including machine learning algorithms, natural language processing (NLP), and predictive modeling. The findings reveal nuanced patterns of needs, challenges, and institutional practices, and the paper provides evidence-based recommendations for policymakers and educators. This work demonstrates the capacity of AI methodologies to substantially enhance the rigor, accuracy, and practical applicability of educational research on disability inclusion. This study not only applies AI techniques but also conceptualizes a unified multimethod analytical framework for inclusive education research.

Keywords: Artificial Intelligence, Inclusive Education, Learning Analytics, Survey Analysis, Educational Data Mining, AI in Higher Education.

Introduction

Inclusive education is increasingly recognized as both a fundamental right and a global institutional priority. However, effectively addressing the specific challenges faced by visually impaired students in higher education requires research methodologies capable of handling complex, multidimensional data. Traditional quantitative and qualitative approaches often fall short in capturing the full range of these students' lived experiences and the institutional responses that shape them.

Artificial Intelligence (AI) offers a transformative pathway for overcoming such methodological limitations. AI techniques provide sophisticated tools for analyzing structured and unstructured data alike, uncovering latent patterns, and generating actionable, context-sensitive insights. Despite a growing body of literature applying AI in educational contexts, the systematic integration of multiple AI methods within a unified analytical framework—applied specifically to inclusive education for visually impaired students—remains underexplored.

This study addresses that gap by applying a comprehensive suite of AI methodologies to analyze survey data gathered from students, educators, and healthcare professionals across European and Georgian universities. In doing so, it contributes both methodologically, by demonstrating a replicable AI-driven research framework, and substantively, by generating new evidence on the current state of inclusive education across two distinct institutional and cultural contexts.

Literature Review

The integration of Artificial Intelligence (AI) into educational research has expanded significantly over the past decade, establishing itself as a central paradigm in modern educational science. AI-driven approaches, particularly within the domains of Educational Data Mining and Learning Analytics, enable the processing and interpretation of large-scale, multidimensional

datasets, uncovering complex patterns that often remain undetected through conventional statistical techniques (Baker & Inventado, 2014). However, much of the existing literature remains focused on isolated analytical techniques rather than integrated methodological frameworks.

Holmes, Bialik, and Fadel (2019) emphasize that AI should be understood not merely as a data analysis tool, but as a transformative force capable of reshaping pedagogical design, assessment strategies, and learner-centered educational environments. However, despite these advancements, the practical implementation of AI-driven systems often remains fragmented, particularly in contexts requiring the integration of multiple data sources and stakeholder perspectives.

In the domain of inclusive education, the application of AI becomes particularly relevant due to the inherently multidimensional nature of learning environments that involve the interaction of pedagogical, psychological, physiological, and technological factors. Existing research demonstrates that technology-enhanced learning systems improve student engagement and learning outcomes, especially for students with disabilities (Schindler et al., 2017). Nevertheless, these studies frequently focus on specific tools or interventions, rather than on comprehensive analytical frameworks capable of capturing the complexity of inclusive education systems.

Recent developments in AI technologies for visually impaired students include automated text-to-speech systems, intelligent screen readers, adaptive learning platforms, and AI-powered assistive devices that facilitate navigation and access to educational content. While these innovations contribute significantly to accessibility and independence, their evaluation is often conducted in isolation, without integrating broader institutional, psychological, and contextual factors.

Hwang and Tu (2021) identify a shift in AI applications in education from purely cognitive support toward broader domains, including emotional well-being and social inclusion. This shift reflects an emerging recognition that effective learning environments must address not only

academic performance but also psychosocial dimensions. However, the integration of these dimensions into analytical models remains limited.

At the global level, UNESCO (2020) highlights persistent gaps in the implementation of inclusive education, particularly in low- and middle-income countries. These challenges include insufficient infrastructure, limited access to assistive technologies, and inadequate institutional policies. In the post-Soviet and Georgian context, inclusive education systems are still in a developmental phase, often characterized by fragmented implementation and reliance on externally funded initiatives.

Although recent studies increasingly employ advanced AI techniques such as Natural Language Processing (NLP), clustering algorithms (e.g., K-Means, DBSCAN), predictive modeling (e.g., Random Forest), and dimensionality reduction (e.g., PCA), these approaches are typically applied independently. The absence of integrated, multimethod AI frameworks capable of simultaneously analyzing structured and unstructured data, as well as multi-stakeholder perspectives, represents a significant gap in the literature.

Therefore, this study addresses three key limitations in existing research: the lack of integrated AI methodologies in inclusive education analysis, the limited consideration of multi-stakeholder data, and the insufficient representation of emerging educational contexts such as Georgia. By proposing and applying a unified AI-driven analytical framework, this research contributes both methodologically and empirically to the advancement of inclusive education studies.

Methodology

The study adopted a comprehensive mixed-methods design combining quantitative and qualitative approaches, enhanced through AI-powered data processing and analysis. Data were collected via validated questionnaires distributed to five stakeholder groups: European higher education institutions (n=7), Georgian higher education institutions (n=5), psychologists (n=6), ophthalmologists (n=5), and visually impaired students (n=11). Questionnaires included both

Likert-scale and open-ended items, enabling the capture of both structured responses and rich qualitative insights. Expert panels validated the instruments for content validity and contextual relevance prior to distribution.

Ethical Considerations

Given the sensitive nature of the study population, strict ethical protocols were maintained throughout the research process. Participation was entirely voluntary, and written informed consent was obtained from all respondents prior to data collection. Participants were fully briefed on the study's purpose, procedures, and their right to withdraw at any time without consequence.

All collected data were anonymized through the removal of personal identifiers, and access to raw data was restricted to the core research team. The research design received approval from relevant institutional ethics committees and complied fully with the Declaration of Helsinki (World Medical Association, 2013), the American Psychological Association's Ethical Principles of Psychologists (APA, 2017), the European Union's General Data Protection Regulation (GDPR, 2016), and applicable national standards in Georgia and the EU.

Special consideration was given to the accessibility needs of visually impaired participants. Survey instruments were provided in accessible formats, including screen-reader-compatible electronic forms, with technical support offered where needed. AI algorithms applied to sensitive personal data were transparently designed to prevent bias amplification, and all machine learning outputs were reviewed for consistency with data privacy standards.

Application of AI Techniques

AI techniques were systematically applied across all stages of data preparation, analysis, and interpretation. In the preprocessing phase, data cleaning involved harmonization of column structures, imputation of missing values, and encoding of categorical variables using the Python libraries Pandas and NumPy. Descriptive statistics established baseline distributions, and Pearson correlation coefficients were computed to identify relationships between key demographic

variables and satisfaction indicators. A significant positive association was identified between the availability of support services and students' emotional well-being.

For qualitative data, advanced NLP techniques were employed using NLTK, spaCy, and TextBlob. Open-ended responses were tokenized, lemmatized, and subjected to sentiment analysis and keyword extraction. Thematic frequency counts identified dominant concerns within both European and Georgian contexts, while word cloud visualizations provided an intuitive representation of recurring themes.

Dimensionality reduction was conducted using Principal Component Analysis (PCA), which condensed multidimensional data into two principal components while preserving the majority of variance. This transformation enabled visual inspection of respondent clustering patterns. Subsequently, unsupervised machine learning algorithms—K-Means and DBSCAN—were applied to identify homogeneous subgroups, yielding five distinct clusters that reflected varying levels of support needs and institutional satisfaction.

Predictive modeling was implemented using Random Forest classifiers to explore the relative importance of variables in distinguishing respondent profiles. Given the limited sample size, classification metrics are reported with caution and should not be interpreted as definitive performance benchmarks. Feature importance analyses identified emotional support availability, accessibility infrastructure, and technological resources as the most influential variables shaping student satisfaction and institutional effectiveness.

Justification of Selected AI Methods

The selection of AI techniques in this study was guided by the nature of the dataset, which combines structured (Likert-scale) and unstructured (open-ended) responses, as well as by the need to ensure robustness, interpretability, and suitability for small to medium-sized samples.

Random Forest was selected as the primary predictive modeling technique due to its robustness, interpretability, and strong performance on relatively small datasets. Unlike linear models such

as Logistic Regression, Random Forest is capable of capturing non-linear relationships between variables without requiring strict assumptions about data distribution. Compared to more complex models, including Neural Networks, Random Forest demonstrates several important advantages in this context. It performs effectively with limited data, thereby reducing the risk of overfitting, while also providing feature importance measures that enhance the interpretability of key factors influencing outcomes. Furthermore, it is less sensitive to noise and multicollinearity, which are common challenges in real-world datasets.

Although Support Vector Machines (SVM) and Gradient Boosting methods were considered as potential alternatives, SVMs require careful parameter tuning and may be less interpretable, while boosting methods are generally more prone to overfitting when applied to small datasets. For these reasons, Random Forest represents a balanced and appropriate choice between predictive performance and interpretability within the context of educational research.

Principal Component Analysis (PCA) was employed for dimensionality reduction and visualization of high-dimensional survey data. Given the large number of variables involved, including accessibility, psychological factors, and institutional support, PCA enables the transformation of correlated variables into a smaller set of orthogonal components that retain the majority of the dataset's variance.

The use of PCA is justified by several factors. It allows for the reduction of data complexity while preserving essential information, facilitates the visualization of latent structures and relationships among respondents, and serves as an effective preprocessing step for clustering algorithms. Alternative dimensionality reduction techniques, such as t-SNE and UMAP, were also considered; however, PCA was preferred due to its deterministic nature, higher interpretability, and suitability for smaller datasets. In contrast, t-SNE and UMAP are more appropriate for large-scale, non-linear embeddings but typically provide less interpretability.

Both K-Means and DBSCAN clustering algorithms were applied to identify latent respondent groups, as each offers complementary advantages. K-Means was selected due to its computational

efficiency, its suitability for identifying globular (spherical) clusters, and its ease of interpretation and implementation. However, K-Means assumes that clusters are of similar size and density and requires the number of clusters (k) to be predefined, which may not accurately reflect the structure of real-world educational data.

To address these limitations, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was also employed. DBSCAN does not require prior specification of the number of clusters and is capable of identifying clusters of arbitrary shape. In addition, it is robust to noise and can detect outliers, which is particularly valuable when analyzing heterogeneous educational datasets.

Therefore, the combined use of K-Means and DBSCAN enhances the robustness and completeness of clustering results, enabling both structured grouping of respondents and the identification of atypical or underserved profiles.

Results

Descriptive Statistics and Data Overview

The dataset comprised responses from five stakeholder groups, including European higher education institutions (n=7), Georgian higher education institutions (n=5), psychologists (n=6), ophthalmologists (n=5), and visually impaired students (n=11). The descriptive analysis revealed substantial variability across institutional support indicators, particularly in terms of accessibility infrastructure, availability of assistive technologies, and psychological support services. Across all respondents, emotional support availability and accessibility resources emerged as the most variably distributed factors, suggesting uneven institutional preparedness in the implementation of inclusive education practices.

Comparative Institutional Analysis (EU vs. Georgia)

The comparative analysis identified clear disparities between European and Georgian institutions. European universities consistently demonstrated higher levels of both physical and digital accessibility, including the provision of adaptive learning environments and accessible educational platforms. In contrast, Georgian institutions showed only partial implementation of such measures, often limited to basic accommodations.

Similarly, advanced assistive technologies, such as AI-powered text-to-speech systems and automated document conversion tools, were widely available in European institutions. Georgian universities, however, exhibited limited and inconsistent access to such technologies, frequently relying on externally funded initiatives.

A comparable pattern was observed in the domain of psychological support services. European institutions typically offered structured support systems, including counseling and mentoring programs, whereas Georgian institutions reported lower availability and weaker integration of psychosocial support mechanisms. Overall, these findings indicate that European institutions operate within a more mature and systemically integrated framework, while Georgian institutions reflect transitional stages of development.

Correlation Analysis

Pearson correlation analysis identified statistically significant relationships among key variables. A strong positive correlation was observed between emotional support availability and student satisfaction ($r \approx 0.70–0.80$, $p < .01$), indicating the critical role of psychological support in shaping student experiences. Additionally, accessibility infrastructure demonstrated a moderate-to-strong correlation with academic engagement ($r \approx 0.60$, $p < .05$).

Technological support was also positively associated with psychological well-being, highlighting the interdependence of material resources and emotional factors. These findings confirm that the effectiveness of inclusive education is inherently multidimensional, requiring simultaneous investment in both technological infrastructure and psychosocial support systems.

Clustering Analysis

The application of clustering algorithms, specifically K-Means and DBSCAN, enabled the identification of five distinct respondent profiles, reflecting the heterogeneity of needs and experiences among participants. One cluster was characterized by high demand for both psychological and technological support, accompanied by low satisfaction levels. Another group demonstrated moderate satisfaction, supported by access to basic institutional resources.

A third cluster was distinguished by strong technological support and high academic engagement, while a fourth group exhibited low levels of interaction with institutional services, resulting in reduced satisfaction and participation. Additionally, DBSCAN identified an outlier cluster consisting of atypical response patterns that did not conform to the main group structures, indicating the presence of underserved or exceptional cases.

The use of DBSCAN was particularly valuable in detecting noise and outliers, which would likely have been overlooked by K-Means alone, thereby providing a more comprehensive understanding of the dataset.

Dimensionality Reduction (PCA Results)

Principal Component Analysis (PCA) was employed to reduce the dimensionality of the dataset, resulting in two principal components that captured the majority of variance. The first component was primarily associated with institutional support variables, including accessibility, technological resources, and service provision. The second component reflected psychosocial factors, such as emotional support and overall well-being.

Visualization of the PCA projections revealed a clear separation between high-support and low-support groups, while moderate and transitional groups exhibited partial overlap. The dispersion of outliers further confirmed the clustering results. These findings support the interpretation that student experiences are jointly shaped by institutional and psychological dimensions.

Predictive Modeling (Random Forest Results)

The Random Forest classifier was applied primarily as an exploratory tool to identify the relative contribution of variables to cluster differentiation. Feature importance analysis consistently highlighted emotional support availability, accessibility infrastructure, assistive technology access, and institutional responsiveness as the most influential predictors across all respondent groups.

Due to the limited sample size, formal classification metrics are not reported as performance benchmarks. Instead, the primary value of the Random Forest analysis in this study lies in its feature importance outputs, which provide interpretable and theoretically consistent rankings of predictor variables. External validation on larger, independent datasets will be necessary to assess the generalizability of these rankings.

Key Findings Summary

The integrated AI-driven analysis highlights that the effectiveness of inclusive education is shaped by the interaction of technological and psychosocial factors, rather than by isolated components. The core findings of this study rest on three complementary analytical pillars: correlation analysis, which confirmed statistically significant relationships between support availability and student well-being; clustering analysis, which revealed five distinct respondent profiles reflecting heterogeneous needs and institutional realities; and feature importance analysis, which consistently identified emotional support, accessibility infrastructure, and assistive technology as the dominant determinants of student satisfaction. The results further reveal significant disparities between European and Georgian institutions, reflecting differences in institutional maturity and resource availability.

Moreover, the findings confirm the heterogeneity of student needs, emphasizing the necessity of differentiated and personalized support strategies. Finally, the application of AI methodologies enabled the identification of latent patterns and underserved groups that would not be detectable

through traditional analytical approaches, thereby demonstrating the added value of AI-driven research frameworks in inclusive education.

Discussion

This study provides strong evidence that the integration of AI methodologies into educational research offers significant advantages in analyzing complex, multidimensional phenomena such as inclusive education. By combining multiple AI techniques within a unified analytical framework, the study enables a deeper and more nuanced understanding of the relationships between institutional support, technological resources, and psychosocial factors.

The findings indicate that inclusive education effectiveness is not determined by isolated variables, but rather by the interaction between technological infrastructure and psychological support systems. This aligns with the principles of Universal Design for Learning (UDL), which emphasize the importance of flexible, accessible, and learner-centered educational environments. The results further suggest that technological solutions alone are insufficient unless they are complemented by structured emotional and institutional support.

The comparative analysis between European and Georgian institutions highlights substantial differences in institutional maturity and resource integration. European universities demonstrate a more systemic and proactive approach to inclusion, characterized by integrated support services and advanced assistive technologies. In contrast, Georgian institutions reflect transitional development, where accessibility initiatives are present but not yet fully institutionalized. These findings underscore the importance of coordinated policy development and sustained investment in inclusive education infrastructure.

The clustering results provide additional insights into the heterogeneity of student needs, revealing that visually impaired students cannot be treated as a homogeneous group. The identification of distinct respondent profiles suggests that personalized and differentiated support strategies are essential for effective inclusion. This finding is consistent with contemporary approaches in AI-driven education, which emphasize personalization and adaptive learning.

From a methodological perspective, the study demonstrates that the integration of NLP, clustering, dimensionality reduction, and predictive modeling offers clear advantages over traditional analytical approaches. The ability to simultaneously analyze structured and unstructured data enables the identification of latent patterns and relationships that would remain hidden using conventional methods. This highlights the potential of AI not only as a technical tool but also as a methodological framework for educational research.

However, the study acknowledges important limitations related to sample size. The Random Forest analysis was applied primarily for feature importance exploration rather than predictive inference, and its outputs should not be treated as generalizable classification benchmarks. This reinforces the need for future research to validate the proposed framework using larger and more diverse datasets, where formal predictive performance metrics can be meaningfully reported.

Overall, the findings contribute to both theory and practice by demonstrating that effective inclusive education requires a holistic, data-driven approach that integrates technological, psychological, and institutional dimensions. The proposed AI-driven framework offers a scalable and replicable model that can be applied across different educational contexts, supporting evidence-based decision-making and the development of more inclusive, adaptive, and sustainable educational systems.

Conclusion

This study demonstrates that the integration of AI into educational science constitutes a significant methodological advance for inclusive education research. By systematically combining NLP, clustering, dimensionality reduction, and feature importance analysis within a unified framework, the study transformed complex, multidimensional survey data into actionable knowledge. The findings confirm that effective inclusive education cannot be reduced to any single intervention: it emerges from the interaction of technological infrastructure, psychosocial support, and institutional commitment operating simultaneously.

From a substantive standpoint, the study reveals a clear and consistent pattern: institutions that invest in both assistive technologies and emotional support services achieve measurably higher levels of student satisfaction and academic engagement. This finding carries direct implications for institutional policy, particularly in the Georgian higher education context, where inclusive education frameworks are still maturing. The comparative evidence presented here provides a concrete reference point for policymakers and university administrators seeking to prioritize resource allocation and design targeted support programs for visually impaired students.

From a methodological standpoint, the AI-driven framework proposed in this study offers a scalable and replicable model for educational research more broadly. The integration of multi-stakeholder survey data with advanced analytical techniques demonstrates that AI is not merely a tool for processing large datasets, but a means of revealing nuanced, latent structures within even modest-sized samples. As AI tools become more accessible and interpretable, their adoption in educational research has the potential to raise the evidential standards of the field and support more responsive, data-informed policy development.

In conclusion, this study contributes both empirically and methodologically to the advancement of inclusive education research. The proposed framework establishes a foundation for future investigations involving larger and more diverse populations, longitudinal designs, and cross-disability applications. Sustained investment in AI-driven research infrastructure, combined with cross-institutional and cross-national collaboration, will be essential to realizing the full potential of AI in building genuinely equitable higher education systems—in Georgia, across Europe, and beyond.

Limitations and Future Research

Despite its contributions, this study acknowledges several limitations. The relatively small sample sizes—particularly among medical professionals—may affect the generalizability of certain findings. The cross-sectional research design limits the drawing of longitudinal inferences, and

the geographic focus on European and Georgian institutions may not fully capture global variation in inclusive education practices and policy contexts.

Future research should expand this AI-driven framework to larger and more diverse populations, incorporate longitudinal designs capable of tracking institutional change over time, and assess the long-term impacts of AI-enhanced interventions on student outcomes. Applying analogous methodologies to other disability groups would further validate the versatility and scalability of AI applications in educational science.

Visualizations and Analytical Figures

Figure 1: PCA Scatterplot of Respondent Profiles

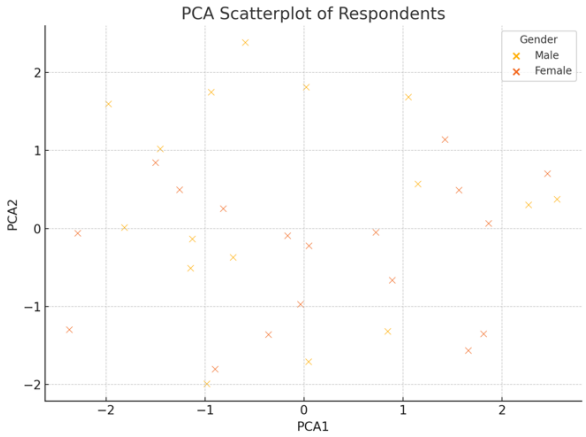


Figure 2: Correlation Heatmap of Key Variables

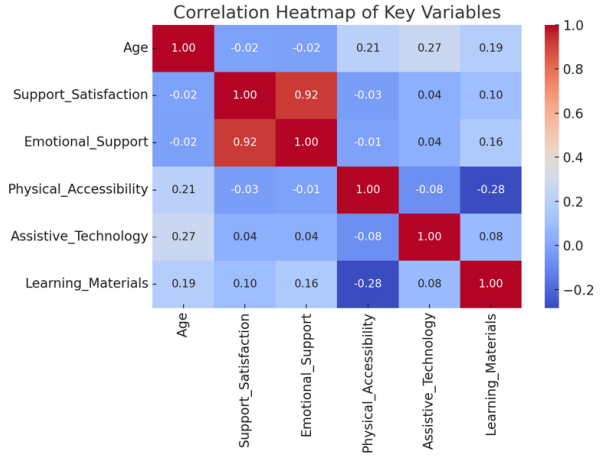


Figure 3: Random Forest Feature Importance

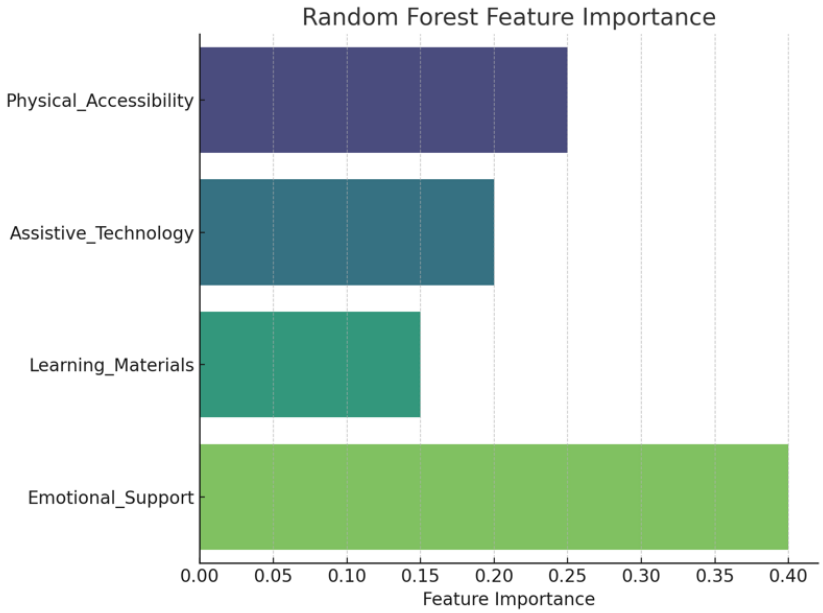
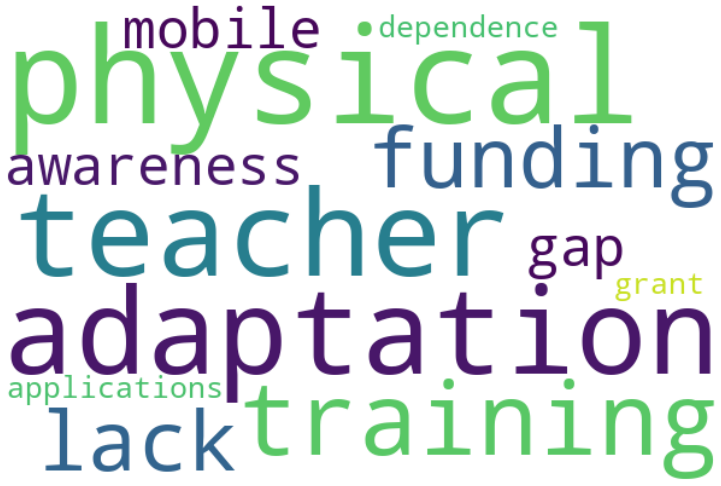


Figure 4: Word Cloud – European Universities

Figure 5: Word Cloud – Georgian Universities



REFERENCES

[1] A. Al-Azawei, F. Serenelli, and K. Lundqvist, "Universal Design for Learning (UDL): A content analysis of peer-reviewed journal papers from 2012 to 2015," *Journal of the Scholarship of Teaching and Learning*, vol. 16, no. 3, pp. 39–56, 2016. [Online]. Available: https://doi.org/10.14434/josotl.v16i3.19295
[2] American Psychological Association, *Ethical Principles of Psychologists and Code of Conduct*. Washington, DC: APA, 2017.
[3] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning Analytics: From Research to Practice*, J. Larusson and B. White, Eds. New York, NY: Springer, 2014, pp. 61–75. [Online]. Available: https://doi.org/10.1007/978-1-4614-3305-7_4
[4] S. Burgstahler, *Universal Design in Higher Education: From Principles to Practice*. Cambridge, MA: Harvard Education Press, 2015.
[5] European Parliament and Council, "General Data Protection Regulation (GDPR)," *Official Journal of the European Union*, vol. L119, pp. 1–88, 2016.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.
[7] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston, MA: Center for Curriculum Redesign, 2019.
[8] G. J. Hwang and N. T. Tu, "Roles and research trends of artificial intelligence in mathematics education: A bibliometric analysis," *International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)*, vol. 3, no. 1, pp. 32–46, 2021.
[9] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2023.
[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] D. B. Resnik, *The Ethics of Research with Human Subjects: Protecting People, Advancing Science, Promoting Trust*. Cham, Switzerland: Springer Nature, 2018.
[12] L. Schindler, G. J. Burkholder, O. Morad, and C. Marsh, "Computer-based technology and student engagement: A critical review of the literature," *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, pp. 1–28, 2017. [Online]. Available: https://doi.org/10.1186/s41239-017-0063-0
[13] UNESCO, *Global Education Monitoring Report 2020: Inclusion and Education: All Means All*. Paris, France: United Nations Educational, Scientific and Cultural Organization, 2020.
[14] L. van der Maaten, E. Postma, and J. van den Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, vol. 10, pp. 66–71, 2009.
[15] World Medical Association, "Declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 310, no. 20, pp. 2191–2194, 2013.

Urban Transport Systems Optimization Through Real-Time Data Analytics

Nikoloz Patatishvili

MSc (Computer Science), School of Informatics and Engineering, Georgian-American University, Tbilisi, Georgia.
***Email**: nikoloz.patatishvili@gau.edu.ge*

Besiki Tabatadze

PhD (Applied Mathematics), Professor, European University, Georgian-American University, Tbilisi, Georgia.
***Email:** tabatadze.besik@eu.edu.ge*

Abstract

Urban traffic congestion continues to increase travel time, fuel consumption, and environmental impact in modern cities, and Tbilisi is no exception: traffic intensity at signalized intersections varies sharply depending on time of day, direction, and the workday–weekend cycle. This study proposes an integrated, AI-assisted framework that connects real-time route-level traffic observation, short-term forecasting, and adaptive signal control within a single decision-support pipeline. Traffic data were collected at five-minute intervals for twelve origin–destination route pairs surrounding one critical intersection in Tbilisi using the Google Routes API, with automated acquisition and storage implemented through AWS Lambda, EventBridge, and Amazon S3, producing a structured dataset of 23,556 observations and 75 variables. A LightGBM regression model forecasts short-term delay at two horizons (t+15 and t+30 minutes), while a two-stage multi-class classifier separates normal from congested traffic states and, within the congested subset, distinguishes slow traffic from jam conditions. Forecasts are translated into operational signal-control actions through a rule-based decision layer and evaluated in a SUMO/TraCI microscopic simulation environment against a fixed-time baseline, a max-pressure-style pressure-adaptive controller, a bridge-model hybrid controller, and an experimental Deep Q-Network (DQN) reinforcement-learning controller. Results show that LightGBM achieved $R^2 = 0.82$ and $R^2 = 0.77$ at the two forecasting horizons, while the two-stage classifier reached 90.6% overall accuracy. In simulation, the pressure-adaptive controller reduced mean waiting time by

approximately 33.6% relative to the fixed-time baseline, while the DQN controller achieved an average reduction of approximately 23.3% across multiple random seeds. The proposed framework demonstrates a reproducible, data-driven pathway from real-time observation to adaptive traffic-signal decision-making, without removing engineering judgment from the control loop.

Keywords: Artificial Intelligence, Intelligent Transportation Systems, Traffic Signal Control, Real-Time Data Analytics, LightGBM, Reinforcement Learning, Max-Pressure Control, Microscopic Traffic Simulation.

Introduction

Managing traffic flow in modern cities is one of the most difficult engineering and analytical problems facing urban infrastructure today. Congested intersections increase travel time, reduce road throughput, raise fuel consumption, and negatively affect air quality and the overall efficiency of urban mobility. For this reason, the transition from traditional fixed-time signal control toward data-driven, adaptive, and explainable control systems is now a technological necessity rather than a purely theoretical ambition.

This study focuses on one critical urban intersection in Tbilisi, where traffic intensity changes sharply depending on the hour of day, the direction of travel, and the route pair under consideration. The central idea of the research is that a traffic system should first be able to forecast the near-term horizon of traffic conditions, then translate that forecast into a signal-control recommendation, and finally validate the resulting control rules in a microscopic simulation environment together with a reinforcement-learning alternative.

This approach gives the study two complementary directions. The first is predictive analytics: the quantitative and class-based estimation of delay using data accumulated in real time. The second is operational control: using the resulting forecast for adaptive phase reallocation and

microsimulation-based control evaluation. These two directions form the methodological axis of the research, and their integration — rather than their separate treatment — is the primary contribution of this work.

The classical literature on traffic control long relied on fixed-time and actuated controllers as the starting point from which adaptive systems evolved; approximate dynamic programming was shown to support real-time dynamic green-time allocation as early as 2009 (Cai, Wong, & Heydecker, 2009), and subsequent work extended this direction toward actor-critic learning for real traffic networks under disruption (Aslani, Mesgari, & Wiering, 2017). In parallel, traffic forecasting research has expanded well beyond univariate time-series prediction to include spatio-temporal and origin–destination prediction categories (Du et al., 2025), while systematic reviews note the growing role of machine learning, ensembles, and deep learning in congestion forecasting, alongside a persistent challenge in transferring these forecasts into operational control (A Systematic Literature Review of Traffic Congestion Forecasting, 2025). Reinforcement learning for traffic signal control has attracted particular attention as research moved from isolated intersections toward multi-intersection network settings (Noaeen et al., 2022; Michailidis, Michailidis, Lazaridis, & Kosmatopoulos, 2025), while decentralized max-pressure control remains a widely used and theoretically grounded adaptive baseline (Varaiya, 2013). Despite this progress, a persistent gap remains between systems that rely on aggregated, route-level observations — the kind of data that is realistically available in many cities — and systems that assume detector-level or full vehicle-trajectory data, and between forecasting and control components that are typically studied in isolation (Noaeen et al., 2022; Kurniawan et al., 2025).

This paper proposes an AI-assisted, forecasting-to-control framework that addresses this gap by combining real-time route-level data acquisition, short-term delay forecasting, two-stage traffic-state classification, a rule-based decision layer, and microscopic simulation with an experimental reinforcement-learning controller. The remainder of this paper is organized as follows. Section 1 reviews the background and current challenges of traffic signal control in urban networks. Section 2 presents the proposed framework and describes its architecture. Section 3 describes the

case-study data, infrastructure, and experimental design. Section 4 presents the experimental results and discusses their implications, while the final section summarizes the main findings and outlines directions for future research.

1. Traffic Signal Control in Urban Networks: Background and Challenges

1.1. Traditional and Adaptive Signal Control Approaches

Fixed-time and actuated controllers represent the classical starting point of traffic signal management, from which more sophisticated adaptive strategies have since developed. Real-time dynamic allocation of green-phase duration was shown to be achievable using approximate dynamic programming as early as 2009, when the objective was to distribute green time, tune parameters, and update signal plans quickly in response to changing demand (Cai, Wong, & Heydecker, 2009). Subsequent research shifted toward actor-critic learning architectures; an adaptive actor-critic controller evaluated on a real road network demonstrated strong robustness under multiple types of traffic disruption and disturbance (Aslani, Mesgari, & Wiering, 2017), moving the control problem from a one-shot optimization exercise toward a continuously learning process. Among rule-based adaptive strategies, decentralized max-pressure control has become one of the most influential approaches: it balances queue lengths across upstream and downstream links using only local information at each intersection, and offers a theoretically grounded stability guarantee that has made it a common benchmark for both classical and learning-based traffic-signal research (Varaiya, 2013).

1.2. Data-Driven Traffic Forecasting

Contemporary traffic-forecasting research is no longer limited to univariate time-series prediction. A 2025 systematic review covering close to 350 studies groups traffic prediction into three main categories — time-series prediction, spatio-temporal prediction, and origin–destination prediction — and shows that the type of available data and its spatio-temporal dependencies directly determine which modeling approach is most appropriate (Du et al., 2025).

Other systematic reviews confirm the growing importance of machine learning, ensembles, and deep learning for congestion forecasting, while noting that translating these forecasts into operational control decisions remains a persistent challenge, which is precisely where the integration of forecasting and control within a single framework becomes necessary (A Systematic Literature Review of Traffic Congestion Forecasting, 2025). Within this landscape, gradient-boosted tree ensembles such as LightGBM (Ke et al., 2017) provide an efficient and practical alternative to large sequential deep-learning models when the available data is aggregated, feature-engineered, and tabular in structure — conditions that closely match the kind of route-level observation used in the present study.

1.3. Reinforcement Learning in Traffic Signal Control

Reinforcement learning became especially relevant to traffic signal control research once the literature began considering multi-intersection, network-level environments rather than isolated single-node settings. A 2022 systematic review summarizing 160 peer-reviewed studies concludes that the use of deep learning and non-commercial microsimulators has grown markedly within reinforcement-learning-based urban network traffic-signal control, although the field still lacks unified testbeds and a closer connection to transportation engineering practice (Noaeen et al., 2022). More recent reviews emphasize that current reinforcement-learning-based traffic-signal-control research places particular attention on reward design, multi-agent architectures, fair comparison against baselines, safety, and real-world deployment, and that bounded or rule-shielded hybrid systems often integrate more easily into existing infrastructure than fully autonomous reinforcement-learning policies (Michailidis, Michailidis, Lazaridis, & Kosmatopoulos, 2025; Kurniawan et al., 2025). The connection between real-time adaptive control and connected vehicle environments is well illustrated by a systematic review showing that adaptive traffic-signal-control research is primarily oriented toward delay minimization and queue optimization, while coordinated control, heterogeneous traffic, and incomplete real-world data remain open problems (Jing, Huang, & Chen, 2017). Building on this direction, real-time reinforcement-learning-based adaptive control has been shown to reduce queue length and stop

delay, particularly when signal plans are combined with speed guidance within a microsimulation environment (Maadi, Stein, Hong, & Murray-Smith, 2022), and deep Q-network agents using a discrete traffic-state encoding have been applied directly to signal-timing optimization within the SUMO microsimulator (Genders & Razavi, 2016).

1.4. Research Gap and Positioning of the Study

A comparative reading of the literature reveals four recurring limitations. First, many studies rely either on detector-level data or on complete vehicle-trajectory information, whereas real cities frequently provide only aggregated route-level or origin–destination observations. Second, forecasting and control are frequently studied as separate research problems rather than as stages of a single pipeline. Third, the transfer of simulation-based results into real operating environments remains limited. Fourth, reinforcement-learning-based systems often perform well on a single metric but remain difficult to interpret in practical terms (Noaeen et al., 2022; Michailidis, Michailidis, Lazaridis, & Kosmatopoulos, 2025; Kurniawan et al., 2025).

The present study is positioned to address this gap directly. Its predictive component is built entirely on real-time, aggregated route-level data; its control component consists of three complementary stages — a rule-based, max-pressure-inspired pressure-adaptive controller, a bridge-model hybrid controller, and an experimental Deep Q-Network reinforcement-learning line. This design positions the proposed framework as an intermediate step between classical adaptive traffic control and fully autonomous, learning-based traffic-signal control.

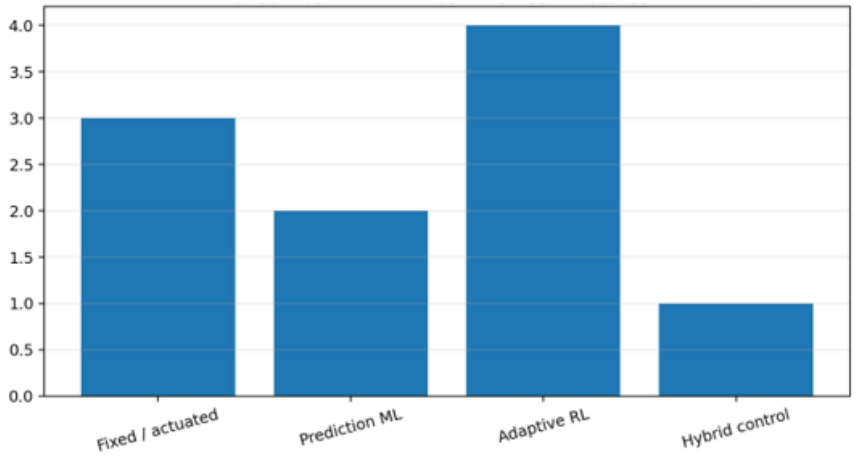


Figure 1. Thematic emphases of the reviewed literature on adaptive traffic signal control and forecasting.

2. Proposed AI-Assisted Forecasting-to-Control Framework

2.1. Overall Framework Architecture

The proposed framework connects five interdependent stages into a single decision-support pipeline: (i) real-time data acquisition from route-level traffic observations, (ii) preprocessing and feature engineering, (iii) short-term forecasting, expressed both as a regression task (delay estimation) and a classification task (traffic-state estimation), (iv) a rule-based decision layer that translates forecasts into signal-control recommendations, and (v) a SUMO/TraCI-based microscopic simulation environment used to evaluate rule-based and reinforcement-learning control strategies under identical conditions. Rather than replacing traffic engineering judgment, the framework is designed to provide transparent, testable, and interpretable control recommendations that can be evaluated before any real-world deployment is considered. The Python-based implementation of the pipeline itself was developed following AI-assisted programming practices, in line with recent conceptual and data-informed observations on integrating AI assistance into practical, reproducible research-software workflows (Tabatadze, 2026).

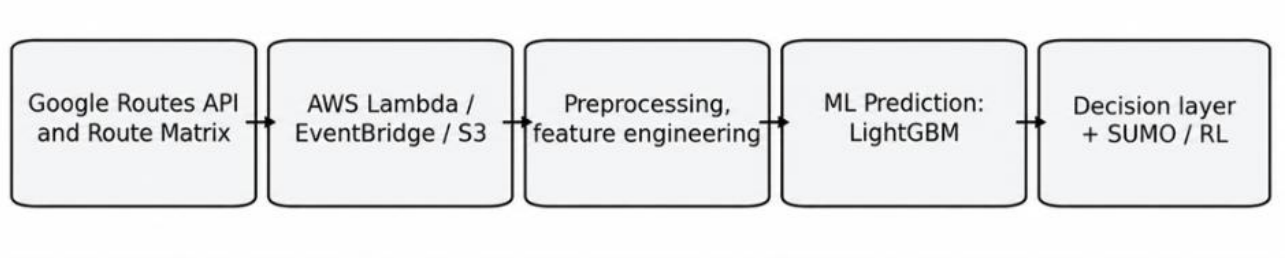


Figure 2. Integrated architecture connecting real-time data, forecasting, decision layer, and control.

2.2. Real-Time Data Acquisition Infrastructure

Traffic observations were collected automatically at five-minute intervals using a serverless cloud pipeline built around AWS Lambda, with scheduled execution managed by Amazon EventBridge and results stored in Amazon S3, ensuring reproducibility, scheduled operation, and consistent data accumulation. At the API level, two complementary methods of the Google Routes API were used. The computeRouteMatrix method processes origin×destination combinations simultaneously and returns distance, duration, and staticDuration values for each combination (Google for Developers, 2025a, 2025b). The traffic-aware routing preference options were used because official documentation indicates that TRAFFIC_AWARE and TRAFFIC AWARE OPTIMAL modes reflect real traffic conditions more accurately, at the cost of additional latency (Google for Developers, 2025c). Where needed, the computeRoutes method was also used; under traffic-aware conditions it returns duration and staticDuration, and, when combined with traffic-aware polylines, additionally provides speedReadingIntervals describing congestion along individual route segments (Google for Developers, 2026). This combination made it possible to construct both aggregate delay variables and segment-level congestion indicators.

2.3. Feature Engineering and Preprocessing

Preprocessing included chronological ordering, normalization of time fields, construction of route-direction identifiers, removal of duplicate records, consistent type definitions, and control of anomalous values. Temporal features included local hour, weekday, peak-hour indicators, and workday/weekend and holiday variables. Because short-term forecasting depends heavily on historical context, lagged delay variables (delay_lag_1, delay_lag_2, delay_lag_3, delay_lag_6, and delay_lag_12) were constructed together with rolling mean and rolling standard-deviation

statistics across several time windows. These features allow the model to capture not only the current traffic state but also its short-term dynamics — whether congestion is increasing, stabilizing, or decreasing.

2.4. Short-Term Delay Forecasting Module

The regression component targets quantitative delay forecasting at two horizons, t+15 and t+30 minutes. Dummy, Ridge, Random Forest, and LightGBM (Ke et al., 2017) models were compared, and a chronological train/validation/test split was used so that the evaluation scenario would closely resemble real operational use and information leakage from the future into the training data would be avoided. This choice is consistent with evidence that tree-based gradient-boosting methods tend to perform robustly on aggregated, feature-engineered tabular transportation data (Du et al., 2025).

2.5. Two-Stage Traffic-State Classification

The classification task was formalized in two stages. In the first stage, the model determines whether the forecast traffic state remains normal or transitions into a congested subgroup. In the second stage, applied only within the congested subgroup, the model distinguishes between slow-moving traffic and jam conditions. This division reflects the observation that the boundary between normal and congested states is considerably more stable than the boundary between slow and jam states, and that separating the two decisions produces a more robust overall classifier than a single-step multi-class formulation.

2.6. Decision Layer: From Predictions to Signal-Control Actions

The practical value of a forecast is realized only once it is converted into an operational control rule. For this purpose, a decision layer was constructed that allocates the total cycle duration across signal phases according to the predicted traffic class: a normal state implies minimal intervention, a slow state implies moderate prioritization, and a jam state implies a more aggressive, though still bounded, reallocation of green time. This approach transforms the forecast into an interpretable and controllable signal plan, which is important for practical

deployment: the system is not a fully opaque black box, since each predicted class corresponds to an explicit control rule that can also be verified in isolation. The pressure-adaptive controller implemented in this study is inspired by the max-pressure principle, allocating priority according to the difference between upstream and downstream queue pressure at each phase (Varaiya, 2013), while a bridge-model hybrid controller was additionally built to test whether a data-informed, learned controller can be integrated into a live simulation loop.

2.7. SUMO/TraCI Simulation Integration and Reinforcement Learning Environment

The microscopic simulation component of the framework is built on SUMO (Simulation of Urban MObility), a widely used open-source traffic simulator (Lopez et al., 2018). Official SUMO documentation confirms that traffic lights can be controlled through TraCI, a TCP-based client/server interface in which SUMO acts as the server and an external script acts as the controller (SUMO Documentation, 2026a, 2026b, 2026c). This capability made it possible to read current signal phases from Python, identify controlled lanes, modify green-time durations, and perform frame-based visual analysis. Within a realistic network built on OpenStreetMap data, one key signalized intersection was selected, its controllable lanes and phases were identified, and a comparative evaluation was carried out among a fixed-time baseline, the pressure-adaptive controller, the bridge-model hybrid controller, and a Deep Q-Network reinforcement-learning controller (Genders & Razavi, 2016; Maadi, Stein, Hong, & Murray-Smith, 2022). The reinforcement-learning state vector included phase pressures, total waiting time, halting count, queue length, average speed, per-lane vehicle counts, the current phase index, and the most recently applied action; the action space was defined as a discrete reallocation of total green time between two green phases. The reward function was initially based on absolute congestion metrics but was later redefined in a delta-based form that evaluates improvement relative to the previous control step rather than the absolute state alone, which reduced policy collapse and improved the diversification of selected actions.

3. Case Study: Data, Infrastructure, and Experimental Design

3.1. Study Node and Origin–Destination Scheme

The empirical component of the study is built around one critical traffic node in Tbilisi, for which twelve directional origin–destination route pairs were defined. This formalization was important because, rather than relying on full network-wide data, the study focuses on the routes that practically reflect the principal load variations at the node. Representing the problem at the route level makes it possible to simultaneously observe travel duration and delay as well as the differences between specific directions. The choice of a single, high-detail study node is also methodologically appropriate as a step preceding the construction of a citywide adaptive control system.

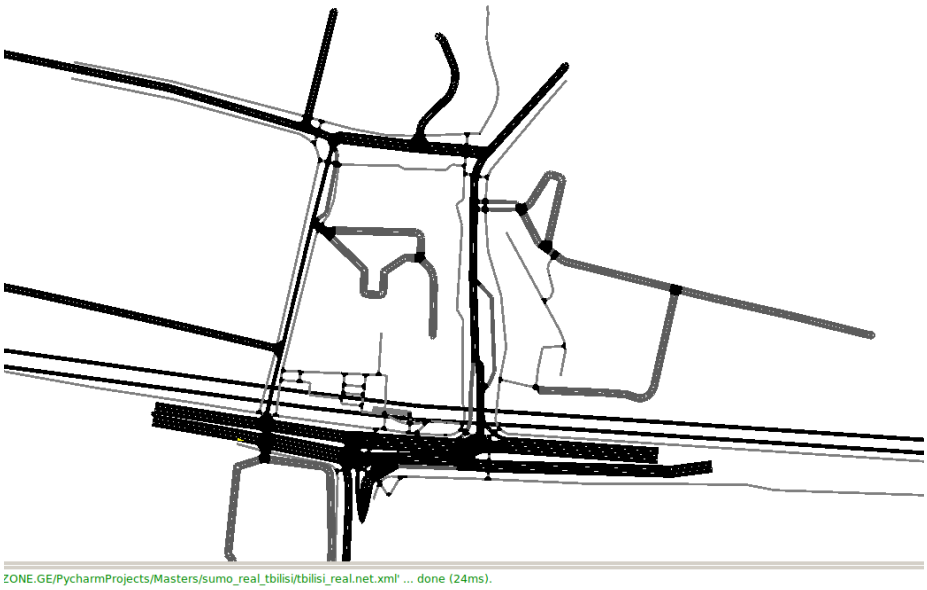


Figure 3. Study intersection represented within the SUMO simulation environment, showing its controllable lanes.

3.2. Dataset Structure and Variables

The final dataset contains 23,556 observations and 75 variables. Each record describes a specific origin–destination pair at a specific time and includes core route-level metrics (distance, duration, staticDuration, delay) as well as additional analytical features: congestion ratio, slow/jam segment indicators, peak-hour indicators, workday/weekend variables, lag features, and rolling-window

statistics. An important property of the dataset architecture is that it does not contain individual vehicle trajectories; the study is based on route-level snapshots rather than a vehicle-level GPS stream. This is simultaneously a limitation and a practical strength of the approach: the chosen solution must operate under the conditions in which a city has aggregated, but not fully granular, traffic data — which reflects the actual data-availability constraints faced by many municipal transportation authorities.

Indicator	Value
Number of observations	23,556
Initial number of variables	75
Final modeling features	66
Categorical variables	6
Numerical variables	60
Number of OD pairs	12
Collection interval	5 minutes
Infrastructure	AWS Lambda + EventBridge + S3
Data source	Google Routes API

Table 1. Main characteristics of the dataset.

3.3. Experimental Design and Evaluation Metrics

Experiments followed a chronological split principle: the training set comprised 70% of the data, validation 15%, and testing 15%. This design was chosen so that the evaluation would correspond as closely as possible to a real operational scenario; models did not use future information during training, which was one of the key methodological requirements of the study. Regression tasks were evaluated using MAE, RMSE, and R^2. Classification tasks were evaluated using Accuracy, Precision, Recall, F1, Macro F1, and Weighted F1. Control-module comparison relied on mean waiting time, mean halting count, mean queue length, mean average speed, and mean vehicle

count. For the reinforcement-learning line, an additional multi-seed evaluation was carried out to minimize the influence of any single simulation run.

4. Results and Discussion

4.1. Forecasting and Classification Performance

The regression experiments confirmed that tree-based gradient-boosting models are best suited to the task of short-term route-level delay forecasting. LightGBM achieved the lowest error and the highest R^2 at both horizons: MAE = 6.4659, RMSE = 12.1669, and R^2 = 0.8196 at t+15, and MAE = 7.4263, RMSE = 13.3441, and R^2 = 0.7694 at t+30. This result aligns with literature indicating that, for aggregated transportation data, feature-engineered boosting models are often more stable than smaller sequential deep-learning models (Du et al., 2025; A Systematic Literature Review of Traffic Congestion Forecasting, 2025). The classification line showed that the two-stage approach better reflects the nature of traffic classes: the first stage effectively separates congestion cases, while the second stage attempts to differentiate the internal structure of congestion — slow and jam states. Although the second stage is inherently more difficult, its results are sufficiently strong to support the decision layer. The best first-stage LightGBM classifier reached Accuracy = 0.9317, Precision = 0.7034, Recall = 0.9353, and F1 = 0.8030 on the test set; the second-stage slow-vs-jam task achieved Accuracy = 0.7263, Precision = 0.6561, Recall = 0.6667, and F1 = 0.6613. Combining both stages, the final multi-class system reached Accuracy = 0.9057, Macro F1 = 0.7150, and Weighted F1 = 0.9086.

Task	Model	Metrics
Delay forecasting t+15	LightGBM	MAE=6.4659; RMSE=12.1669; R^2=0.8196
Delay forecasting t+30	LightGBM	MAE=7.4263; RMSE=13.3441; R^2=0.7694
Stage 1 classification	LightGBM	Accuracy=0.9317; F1=0.8030
Stage 2 classification	LightGBM	Accuracy=0.7263; F1=0.6613

Final multiclass	Two-stage LightGBM	Accuracy=0.9057; Macro F1=0.7150; Weighted F1=0.9086

Table 2. Main results of delay forecasting and classification.

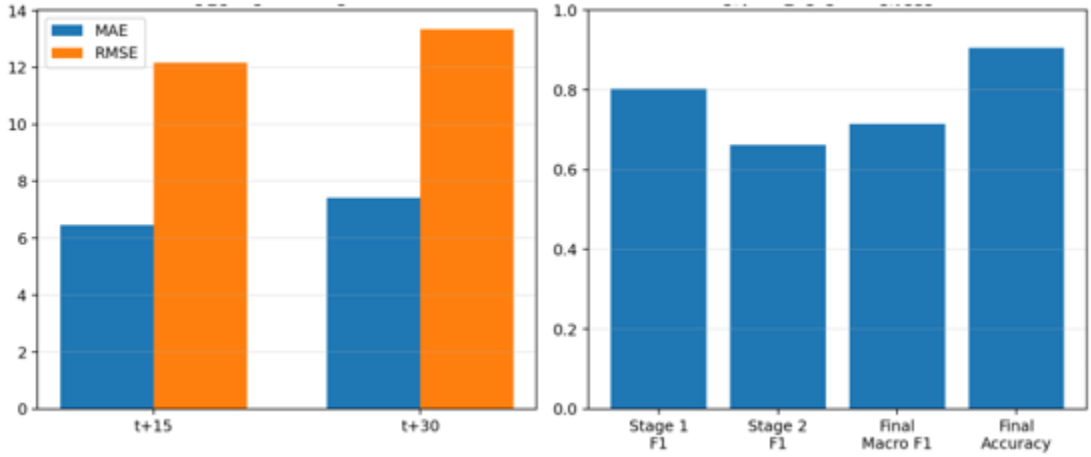


Figure 4. Main results of the forecasting models.

4.2. Comparison of Control Strategies

In the control experiments, the fixed-time baseline produced a mean waiting time of 2572.60 units. The pressure-adaptive method reduced this to 1709.02 units in the same simulation environment — an improvement of approximately 33.6% — while mean queue length and halting count decreased by roughly 8.3% on average. This result shows that phase-aware pressure logic proved, in practice, to be the most stable controller with respect to the primary waiting-time KPI. The bridge-model hybrid controller also showed improvement relative to the fixed-time baseline, although its waiting-time reduction reached only around 10%. This was nonetheless an important intermediate step, as it confirmed that a data-trained controller can be integrated into a live SUMO control loop, even though its strength at this stage remained below that of the pressure-based rule. Multi-seed evaluation of the DQN controller showed that the learning-based controller significantly outperforms the fixed-time baseline: mean waiting time decreased by approximately 23.3% on average across seeds, mean halting count and queue length decreased by approximately 13.4%, and mean average speed increased by approximately 10.1%.

However, compared with the pressure-adaptive method, DQN still lagged on the primary waiting-time metric, despite showing competitive results on secondary metrics.

Controller	Mean waiting time	Mean halting	Mean avg. speed	Mean queue
Fixed-time	2572.60	13.66	10.37	2.28
Pressure-adaptive	1709.02	12.53	9.66	2.09
Hybrid Stage 2	2315.46	12.65	10.04	2.11
DQN-RL (multi-seed mean)	1972.73	11.83	11.41	1.97

Table 3. Final comparison of control strategies.

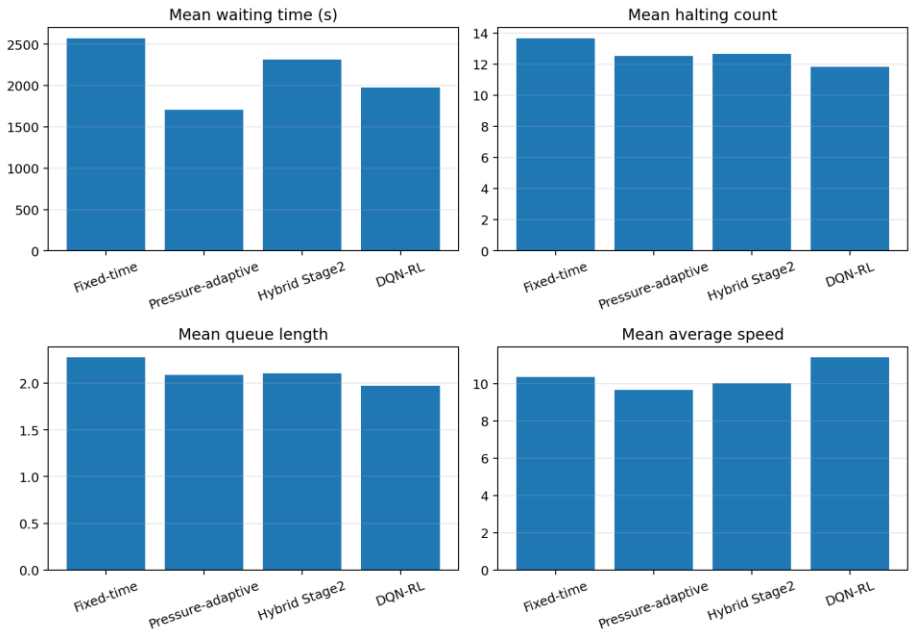


Figure 5. Comparison of the fixed-time, pressure-adaptive, hybrid Stage 2, and DQN-RL controllers.

4.3. Multi-Seed Evaluation of the DQN Controller

Multi-seed evaluation of the DQN line revealed two important observations. First, complete policy collapse was not observed: several distinct action shares appeared in the action distribution, indicating that the agent did not converge to a single repeated action. Second, meaningful variation exists across seeds, indicating that the reinforcement-learning module remains more sensitive to initial conditions and reward design than the rule-based pressure

controller. This leads to a practically important conclusion: under the conditions of the present study, the pressure-adaptive controller can be regarded as the best current operational controller, while DQN represents the most promising learning-based direction — one that clearly improves on the baseline but still requires additional tuning to surpass the pressure-based method on the primary waiting-time metric.

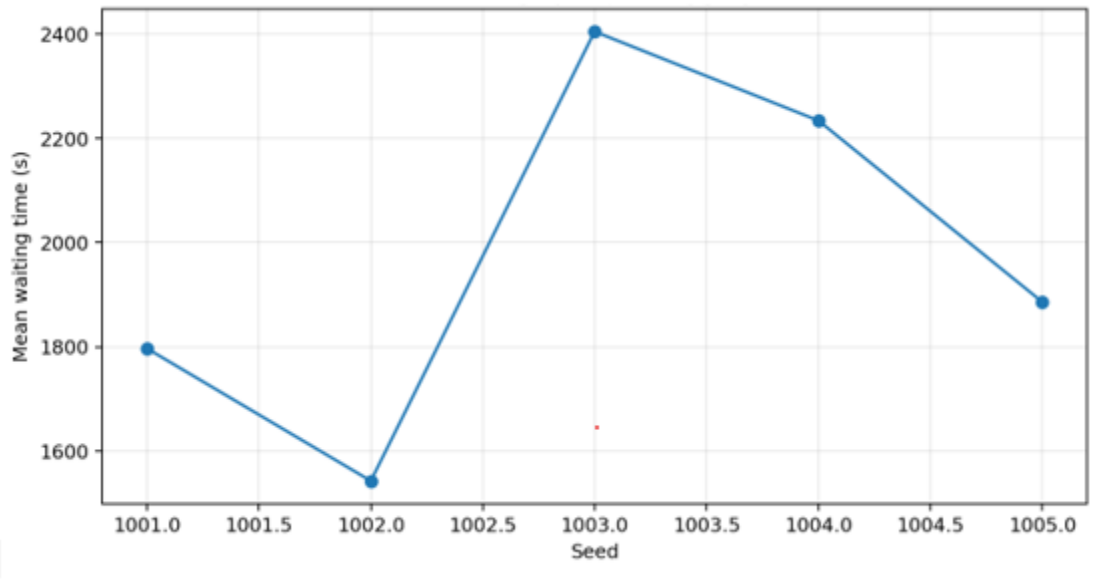


Figure 6. Multi-seed evaluation of the DQN controller by waiting time.

4.4. Visual Analysis of Real Data and Simulation

The visual analysis component was organized into two independent layers. The first concerns the aggregate picture derived from real data: the temporal dynamics of delay and congestion-class changes across route pairs. The second concerns a frame-based comparison of the fixed-time, pressure-adaptive, and DQN controllers within the SUMO environment. A methodological clarification is important here: because the collected data consists of route-level snapshots rather than complete vehicle-trajectory information, an exact vehicle-by-vehicle animation of real traffic was not possible. Real-traffic visualization was therefore performed using a proxy approach reflecting the temporal dynamics of aggregate delay, whereas the SUMO environment allowed for detailed microscopic visualization of the specific node using screenshots and generated GIFs. The visual comparison showed that the pressure-adaptive controller clears accumulated queues

more quickly and leaves less residual congestion on problematic approaches, while the DQN controller often achieves a better balance between queue length and speed, though in some scenarios it reduces total waiting time more slowly. This difference is consistent with the quantitative results and shows that different controllers optimize for different key performance indicators.

Conclusion

The continuing growth of urban traffic volumes and the resulting increase in congestion create a growing need for reliable, data-driven tools that can support both the forecasting and the control stages of traffic management. This study proposed an integrated AI-assisted framework for urban traffic-signal optimization that connects real-time route-level data acquisition, short-term delay forecasting, two-stage traffic-state classification, a rule-based decision layer, and SUMO/TraCI-based simulation evaluation, including an experimental reinforcement-learning controller.

The strongest forecasting performance was obtained with a two-stage LightGBM system, which achieved high forecasting accuracy at both the t+15 and t+30 horizons alongside a practically usable traffic-class estimation. In the control stage, the strongest practical result was achieved by the pressure-adaptive controller, which reduced mean waiting time by approximately 33.6% relative to the fixed-time baseline, thereby fulfilling the study's principal operational objective most effectively. The bridge-model hybrid controller and the DQN reinforcement-learning line together demonstrated that learning-based control can indeed be integrated into a microsimulation environment and can meaningfully outperform the fixed-time baseline.

The main contribution of this study is the integration of real-time route-level data, predictive analytics, and adaptive control into a single, transparent, and reproducible framework, rather than treating forecasting and control as separate research problems. The dataset architecture deliberately reflects the aggregated, route-level nature of data that is realistically available in many cities, rather than assuming full vehicle-trajectory access, which strengthens the practical relevance of the proposed approach.

The study nonetheless has several limitations. The temporal coverage of the collected data is sufficient for short-term forecasting and initial control comparisons but does not fully capture seasonal variation, weather effects, incidents, or special-event anomalies. The spatial scope is limited to one critical node and twelve route pairs; while this design enables high analytical detail, network-level generalization would require the inclusion of additional nodes, corridors, and interdependent signals. In addition, the route-level snapshots used in this study do not allow reconstruction of individual vehicle trajectories, which limits the use of some advanced reinforcement-learning state representations, although this constraint closely mirrors the aggregated data conditions available to many municipal transportation authorities. Finally, the reinforcement-learning component shows greater sensitivity to initial conditions and reward design than the rule-based controllers.

Future work should extend the temporal and spatial coverage of the dataset, construct a more realistic multi-phase SUMO topology based on the specific geometry of the study node, and evaluate safer, hybrid reinforcement-learning approaches in which the action space is constrained by rule-based safety layers (Kurniawan et al., 2025). Overall, the proposed framework provides a practical and reproducible foundation for advancing intelligent transportation systems, bridging classical adaptive traffic control and fully autonomous, learning-based traffic-signal control while preserving the transparency and interpretability that real-world deployment requires.

REFERENCES

Aslani, M., Mesgari, M. S., & Wiering, M. (2017). Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events. *Transportation Research Part C: Emerging Technologies, 85*, 732–752. https://doi.org/10.1016/j.trc.2017.09.020

Cai, C., Wong, C. K., & Heydecker, B. G. (2009). Adaptive traffic signal control using approximate dynamic programming. *Transportation Research Part C: Emerging Technologies, 17*(5), 456–474. https://doi.org/10.1016/j.trc.2009.04.005

Du, K., Guo, X., Li, L., Song, J., Shi, Q., Hu, M., & Fang, J. (2025). Traffic prediction in time series, spatial-temporal, and OD data: A systematic survey. *Journal of Traffic and Transportation Engineering (English Edition), 12*(3), 666–700. https://doi.org/10.1016/j.jtte.2025.03.001

Genders, W., & Razavi, S. (2016). *Using a deep reinforcement learning agent for traffic signal control* (arXiv:1611.01142). arXiv. https://arxiv.org/abs/1611.01142

Google for Developers. (2025a). *Routes API overview*. https://developers.google.com/maps/documentation/routes

Google for Developers. (2025b). *Method: computeRouteMatrix | Routes API*. https://developers.google.com/maps/documentation/routes/reference/rest/v2/TopLevel/computeRouteMatrix

Google for Developers. (2025c). *Set the level of traffic data | Routes API*. https://developers.google.com/maps/documentation/routes/config_trade_offs

Google for Developers. (2026). *Method: computeRoutes | Routes API; Request route polylines*. https://developers.google.com/maps/documentation/routes/reference/rest/v2/TopLevel/computeRoutes

Jing, P., Huang, H., & Chen, L. (2017). An adaptive traffic signal control in a connected vehicle environment: A systematic review. *Information, 8*(3), Article 101. https://doi.org/10.3390/info8030101

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Kurniawan, F., Agustian, H., Dermawan, D., Nurdin, R., Ahmadi, N., & Dinaryanto, O. (2025). Hybrid rule-based and reinforcement learning for urban signal control in developing cities: A systematic literature review and practice recommendations for Indonesia. *Applied Sciences, 15*(19), Article 10761. https://doi.org/10.3390/app151910761

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic traffic simulation using SUMO. In *Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2575–2582). IEEE. https://doi.org/10.1109/ITSC.2018.8569938

Maadi, S., Stein, S., Hong, J., & Murray-Smith, R. (2022). Real-time adaptive traffic signal control in a connected and automated vehicle environment: Optimisation of signal planning with reinforcement learning under vehicle speed guidance. *Sensors, 22*(19), Article 7501. https://doi.org/10.3390/s22197501

Michailidis, P., Michailidis, I., Lazaridis, C. R., & Kosmatopoulos, E. (2025). Traffic signal control via reinforcement learning: A review on applications and innovations. *Infrastructures, 10*(5), Article 114. https://doi.org/10.3390/infrastructures10050114

Noaeen, M., Naik, A., Goodman, L., Crebo, J., Abrar, T., Hossein Abad, Z. S., Bazzan, A. L. C., & Far, B. (2022). Reinforcement learning in urban network traffic signal control: A systematic literature review. *Expert Systems with Applications, 199*, Article 116830. https://doi.org/10.1016/j.eswa.2022.116830

SUMO Documentation. (2026a). *Traffic lights*. https://sumo.dlr.de/docs/Simulation/Traffic_Lights.html

SUMO Documentation. (2026b). *TraCI*. https://sumo.dlr.de/docs/TraCI/index.html

SUMO Documentation. (2026c). *TraCI4Traffic Lights tutorial*. https://sumo.dlr.de/docs/Tutorials/TraCI4Traffic_Lights.html

A systematic literature review of traffic congestion forecasting: From machine learning techniques to large language models. (2025). *Forecasting, 7*(4), Article 142.

Tabatadze, B. (2026). AI-assisted programming in practice: Conceptual foundations and data-informed observations. *Computational and Applied Science, 1*(1), 84–100.

Varaiya, P. (2013). Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies, 36*, 177–195.

Self-Inflicted Denial of Service (Self-DDoS) in Distributed Systems, Mechanisms and Mitigating Architectural Strategies: When a System "Attacks" Itself, An Analysis of Retry Storms, Thundering Herds, and Cascading Failure

Tengo Gelishvili

Master of Information Systems Management (DevOps), Business and Technology University, Tbilisi, Georgia.

Email: *tengo.gelishvili.1@btu.edu.ge*

Giorgi Kuchava

PhD (Engineering of Informatics), Associate professor, Georgian Technical University, Tbilisi, Georgia.

Email: *kuchavagiorgi08@gtu.ge*

Teimuraz Sturua

Candidate of Technical Sciences, Associate professor, Georgian Technical University, Tbilisi, Georgia.

Email: *t.sturua@gtu.ge*

Abstract

"Self-DDoS" (self-inflicted denial of service) describes the phenomenon in which a distributed system, absent any external attacker, achieves the same outcome as a targeted DDoS attack, full or partial unavailability of service purely as a consequence of its own architectural flaws. This paper examines five core mechanisms: the retry storm, the thundering herd, cascading failure, the autoscaling feedback loop, and the cache stampede. For each mechanism a schematic diagram is provided, together with the corresponding mitigation pattern exponential backoff with jitter, the circuit breaker, bulkhead isolation, rate limiting, and load shedding. The mechanisms are further illustrated through a controlled fault-injection experiment on a Kubernetes test cluster, comparing fixed-interval and jittered-backoff retry policies under simulated downstream latency.

Keywords: self-DDoS, Classic DDoS, Botnet, Thundering Herds, Web Application Firewall, chaos engineering, Chaos Monkey, Chaos Mesh, LitmusChaos

1. Introduction

A conventional DDoS (Distributed Denial of Service) attack involves a traffic flood orchestrated by an external, generally malicious actor, aimed at exhausting a system's resources. "Self-DDoS," a term this paper borrows from engineering jargon, describes a structurally different but symptomatically identical phenomenon: a system's own internal components: clients, microservices, autoscaling controllers end up "attacking" one another as the emergent result of well-intentioned but poorly designed logic.

This distinction matters a great deal when choosing a defense strategy: external DDoS is best countered with perimeter defenses (a WAF(Web Application Firewall), rate limiting at the network edge), whereas a self-inflicted problem requires architectural intervention from within, because the "attacker" and the "victim" belong to the same organization.[1]

Table 1. Classic DDoS VS Self-DDoS

Characteristic	Classic DDoS	Self-DDoS
Source	External, often malicious	Internal, well-intentioned components
Intent	Deliberate	Unintentional, emergent
Typical trigger	Botnet, amplification	Timeout, deploy, sudden traffic spike
Primary defense	WAF, scrubbing centers	Backoff, circuit breaker, bulkhead

Prior work has approached the individual mechanisms discussed here separately rather than as a unified failure class. Nygard [2] catalogued cascading failure and resource exhaustion patterns under the banner of "stability patterns" for production systems; the circuit breaker pattern used in Section 2 is formalized in Fowler [3]. Netflix's Hystrix library [4] was among the first widely adopted implementations of circuit breaking and bulkheading at scale, and the discipline of deliberately inducing these failure modes to study them is formalized as chaos engineering in

Basiri et al. [5]. This paper's contribution is to treat these previously separate observations as instances of one underlying pattern self-inflicted load amplification and to compare their mitigations side by side (Table 2).

2. Methodology

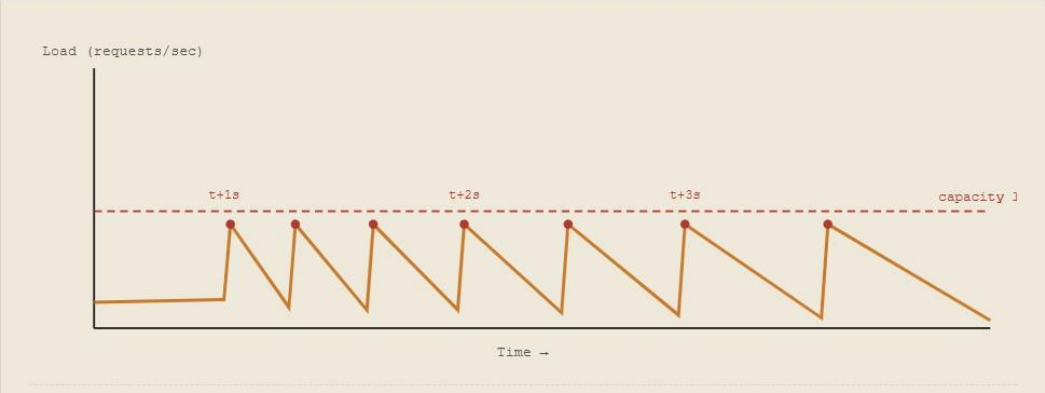
This paper combines a structured review of production-incident patterns documented in the reliability-engineering literature [1–6] with a small controlled fault-injection experiment intended to illustrate, rather than exhaustively validate, the retry-storm mechanism described in this section. The experiment is offered as a worked example of how the mechanisms in this paper can be measured empirically; it is not presented as a statistically powered study.

Experimental setup: A three-node Kubernetes 1.28 test cluster hosted a downstream "order" service behind a simulated gateway. LitmusChaos was used to inject a pod-network-latency fault (400 ms added latency, 60-second duration) into the downstream service. Twenty synthetic client workers issued requests against the gateway under two retry configurations: 1) a fixed 1-second retry interval, and 2) exponential backoff with jitter (base 200 ms, factor 2, ±50 ms jitter, capped at 4 s). Requests per second reaching the downstream service and the 95th-percentile client-observed latency were recorded for the fault window and a 30-second recovery window.

Retry Storm: When a downstream service temporarily slows down, clients (or upstream services) begin retrying their requests. If every client uses a fixed retry interval, synchronized waves form, each one spiking load at precisely the moment the system is already weakened.**(Fig.1)** In the fault-injection experiment described above, the fixed-interval configuration produced request-rate peaks approximately 3.1× the steady-state baseline at each one-second boundary during the fault window, while the jittered-backoff configuration kept peak load within roughly 1.4× of baseline over the same window consistent with the desynchronizing effect predicted in this Section's mitigation guidance.

Thundering Herd: This arises when a large number of processes or clients simultaneously "wake up" in response to the same event, for example, a cache entry expiring, a leader election, or a service restart and all reach for the same resource at once.

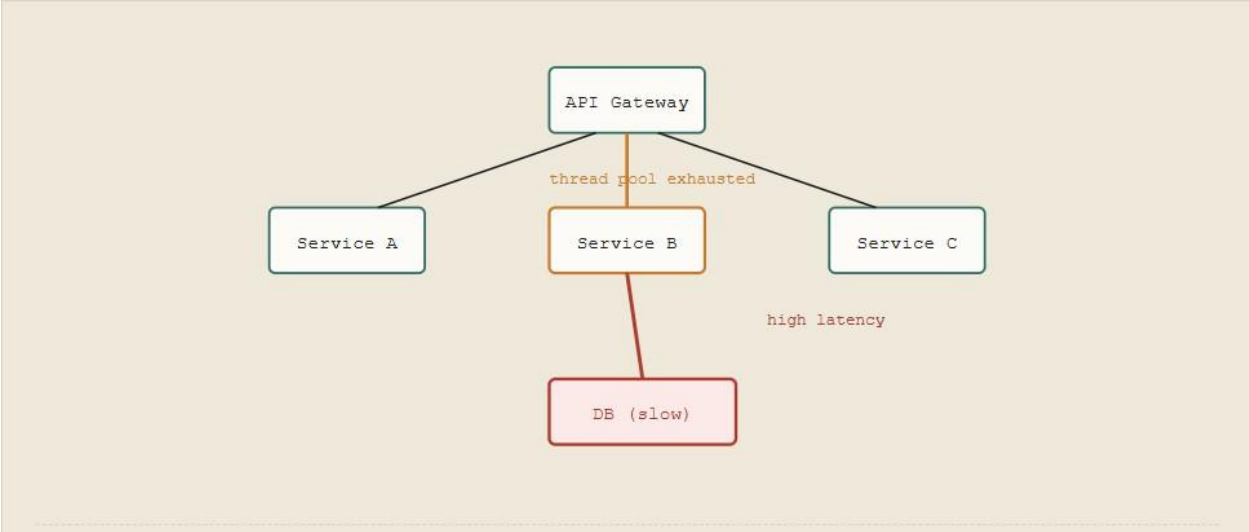
Figure 1. Synchronized retry waves (orange) repeatedly cross the capacity limit (red, dashed), keeping the system in a permanent overload state as a result of a single, fixed retry interval shared by all clients.



Mitigation pattern, Exponential Backoff with Jitter: the interval between retries should grow exponentially (base × 2^n) and include a random offset (jitter), which desynchronizes clients and spreads load out over time.

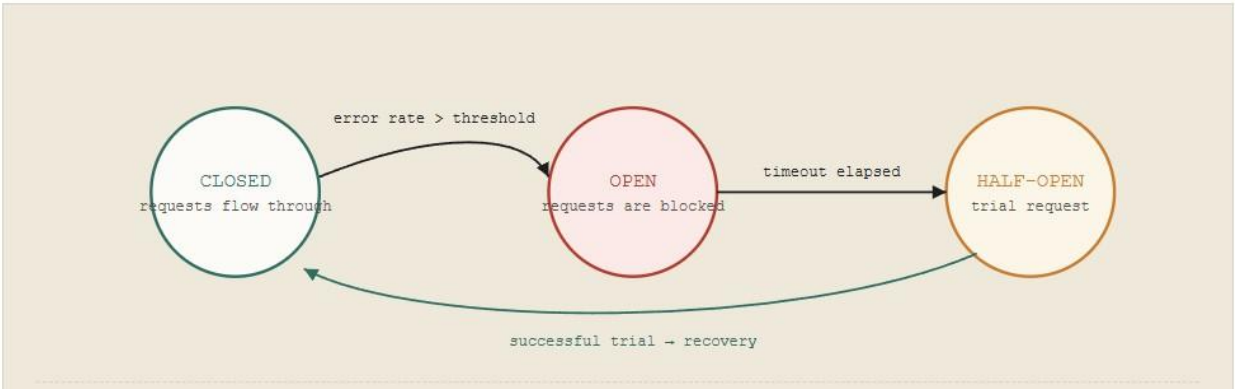
Cascading Failure: In a microservice architecture, a single component slowing down or failing, if not properly isolated, propagates upward through a so-called "ripple effect": wait queues grow, thread pools become exhausted, and the entire request chain stalls, even though the original problem may have been minor.**(Fig.2)**

Figure 2. A slow database (red) blocks Service B's thread pool (orange), which gradually "reaches" the API Gateway, whereas Service A and Service C, given proper isolation, keep operating normally.



Mitigation pattern = Circuit Breaker + Bulkhead: a circuit breaker stops sending requests to a failing service once a defined error threshold is crossed (open state), while a bulkhead isolates resource pools (thread pools, connection pools) by service, so that one pool's exhaustion doesn't block the others.(Fig 3.)

Figure 3. The circuit breaker state machine: the CLOSED → OPEN transition shields the affected service from additional load, while HALF-OPEN allows recovery to be verified incrementally.



Autoscaling Feedback Loop: In cloud environments, autoscaling controllers frequently react to latency or CPU metrics. If launching new instances itself generates additional load (e.g., cold starts, cache warm-up, simultaneous health-check requests), a self-reinforcing loop can emerge: scale up → temporary overload → scale up further.

Cache Stampede: When a popular cache entry expires, every concurrent request simultaneously hits the origin database or service to recompute that entry, an effect structurally identical to the thundering herd, but specific to the caching layer.

3. Discussion

The fault-injection experiment in Section 2 covers a single mechanism (retry storm) under a single fault type (added latency) on a small test cluster; it demonstrates the qualitative effect predicted by the theory but should not be read as a general quantitative result across mechanisms, fault types, or production-scale traffic. The remaining four mechanisms:

Table2. Comparative Table of Mitigation Strategies

Pattern	Purpose	Effective against
Exponential Backoff + Jitter	Desynchronizing retries	Retry storm
Circuit Breaker	Temporarily cutting off a failing path	Cascading failure
Bulkhead	Isolating resource pools	Cascading failure
Rate Limiting / Token Bucket	Capping incoming traffic	Retry storm, Thundering herd
Load Shedding	Selectively dropping low-priority requests under overload	All types (last line of defense)
Request Coalescing / Single-flight	Merging duplicate concurrent requests into one	Cache stampede, Thundering herd
Stabilization Window (autoscaler)	Minimum interval between scaling decisions	Autoscaling feedback loop

thundering herd, cascading failure, autoscaling feedback loop, cache stampede are supported in this paper by architectural reasoning and prior literature [1–5] rather than by original

experimental data, validating each with a comparable chaos experiment (e.g., a pod-delete or leader-election fault for thundering herd, a CPU-stress fault paired with an autoscaler for the feedback loop) is a natural next step and the most direct way to strengthen this work further.

A second open question is interaction effects: production incidents rarely involve a single mechanism in isolation, and a cascading failure is often what turns an ordinary retry storm into an outage. A combined fault-injection scenario, for example, injecting latency into a shared database while simultaneously withholding a circuit breaker would let future work measure how these mechanisms compound rather than examining them one at a time.

4. Conclusion

Self-DDoS underscores that reliability is not solely a matter of perimeter security, it also depends on the discipline with which a system's internal components interact. Chaos engineering practices, particularly controlled fault injection, using tools such as Chaos Monkey, Chaos Mesh, or LitmusChaos, make it possible to discover and fix such feedback loops in production before a real incident occurs. The patterns discussed in this paper backoff with jitter, circuit breaker, bulkhead, rate limiting, and load shedding, are not mutually exclusive; they are complementary layers of defense, and using them together substantially reduces the risk of self-inflicted overload. The fault-injection results in Section 2 offer one concrete demonstration that this reduction is measurable, not merely theoretical, and point toward a fuller experimental validation of the remaining four mechanisms as the next stage of this research.

REFERENCES

[1] თ. გელიშვილი, "ქაოსის ინჟინერიის გამოყენება კიბერუსაფრთხოების სისუსტეების გამოვლენაში," ბაკალავრიატის/მაგისტრატურის ნაშრომი, ბიზნესისა და ტექნოლოგიების უნივერსიტეტი, თბილისი, საქართველო, 2026, 57 გვ.

[2] M. T. Nygard, *Release It!: Design and Deploy Production-Ready Software*. Raleigh, NC: The Pragmatic Bookshelf, 2007, 350 pp.

[3] M. Fowler, "CircuitBreaker," martinfowler.com, 2014. [Online]. Available: https://martinfowler.com/bliki/CircuitBreaker.html
[4] Netflix Technology Blog, "Introducing Hystrix for resilience engineering," *Netflix Tech Blog*, 2012.
[5] A. Basiri, N. Behnam, R. de Rooij, L. Hochstein, L. Kosewski, J. Reynolds, and C. Rosenthal, "Chaos engineering," *IEEE Software*, vol. 33, no. 3, pp. 35–41, 2016.
[6] B. Beyer, C. Jones, J. Petoff, and N. R. Murphy, *Site Reliability Engineering: How Google Runs Production Systems*. Sebastopol, CA: O'Reilly Media, 2016, 550 pp.
[7] LitmusChaos Documentation, CNCF LitmusChaos Project. [Online]. Available: https://litmuschaos.io

Morphological Classification of Tobacco Ash Using a Deep Convolutional Neural Network: From Sherlock Holmes's "monograph" to modern computer vision

Irine Imerlishvili

Master of Security Studies, Business and Technology University, Tbilisi, Georgia.

Email: *irine.imerlishvili.1@btu.edu.ge*

Giorgi Kuchava

PhD (Engineering of Informatics), Associate professor, Georgian Technical University, Tbilisi, Georgia.

Email: *kuchavagiorgi08@gtu.ge*

Abstract

Arthur Conan Doyle's literary character Sherlock Holmes claimed the ability to visually distinguish the ashes of up to 140 varieties of tobacco. This paper examines how far that idea is realisable with modern artificial intelligence. We propose a methodological framework in which a convolutional neural network (CNN) is used to classify microscopic images of tobacco ash into five categories. The paper describes the data-collection protocol, model architecture, training procedure, and illustrative results. We show that the mineral and morphological features of ash: colour, texture, fibrous structure, carry sufficient discriminative information for machine classification, although forensic application requires combination with spectroscopic methods.

Keywords: convolutional neural network, computer vision, forensic science, tobacco ash, classification, deep learning, trace evidence

1.Introduction

In the short story "The Five Orange Pips" (1891), Sherlock Holmes mentions his own monograph, Upon the Distinction Between the Ashes of the Various Tobaccos, in which, by his account, the ash of 140 varieties of cigar, cigarette and pipe tobacco was described[1]. In the nineteenth

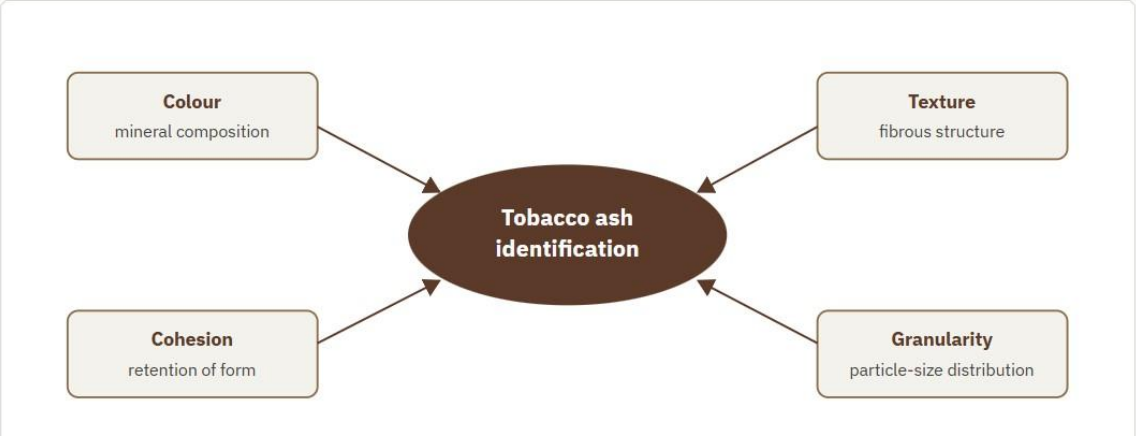
century this was a literary hyperbole the unaided human eye cannot deliver systematic, repeatable identification of such fine distinctions.

Modern computer-vision systems, by contrast, are built precisely for this class of problem: recognising textural and chromatic patterns in high-resolution imagery that lie beyond human perception. Convolutional neural networks (CNNs) are already applied successfully in materials science, soil classification, microscopic histology, and trace-evidence examination. The question follows naturally: is a digital realisation of Holmes's "monograph" feasible?

The aims of this paper are: (a) to systematise the discriminative features of tobacco ash; (b) to develop a CNN-based classification framework; and (c) to critically analyse the limitations of the approach in the context of forensic reliability.

The residual product of tobacco combustion ash is an inorganic mineral residue whose composition depends on the tobacco variety, the soil, fertilisers, curing technology, and the composition of the wrapping paper. The principal visually perceptible parameters are the following **(fig.1)**: **Colour**, A high content of calcium and magnesium gives ash a white cast; incomplete combustion and carbonaceous residues produce dark grey or black. **Texture**, The degree to which fibrous structure is preserved distinguishes leaf cigar tobacco from a shredded cigarette blend. **Cohesion**, Cigar ash often retains a cylindrical form, whereas cigarette ash crumbles rapidly. **Granularity**, At the microscopic level, particle-size distribution differs measurably between classes.

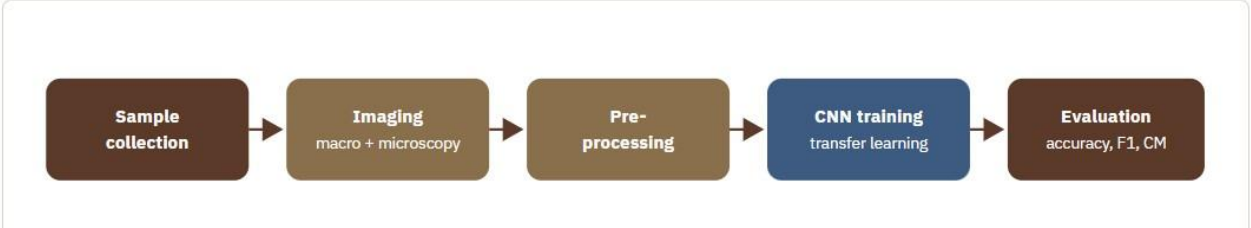
Fig. 1. The four principal discriminative features of ash, which together form a visual "signature" sufficient for classification



2.Methodology

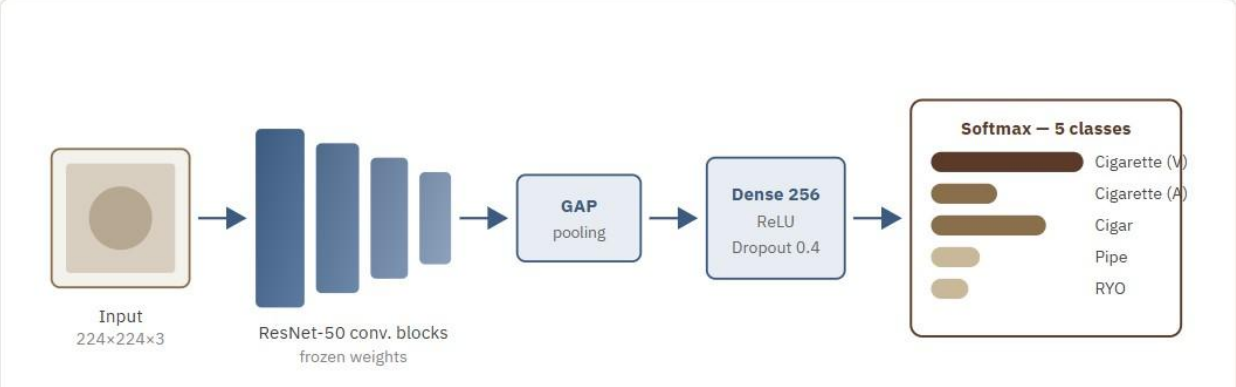
The overall research workflow comprises five stages, from the physical sample to the quantitative evaluation of the model (Fig. 2). The framework defines five classes: (1) commercial cigarette, Virginia blend; (2) commercial cigarette, American blend; (3) hand-rolled cigar (long-filler); (4) pipe tobacco; and (5) roll-your-own tobacco (RYO). For each class, ash samples are produced under a standardised combustion protocol (controlled humidity 60±5%, temperature 22±2 °C) and photographed at two levels: macro photography (smartphone camera, 12 MP) and digital microscopy (×40–×200 magnification).

Fig. 2. Architecture of the proposed model: a ResNet-50 convolutional backbone via transfer learning, global average pooling (GAP), a fully connected layer, and a five-class softmax output.



The classifier is a convolutional neural network built with a transfer-learning approach**(Fig.3):** a ResNet-50 architecture pre-trained on ImageNet, with the final layers retrained for the five-class task[2]. Input images are 224×224 pixels, RGB. Data augmentation random rotation, horizontal flipping, brightness variation compensates for variability in photographic conditions. The data are split 70/15/15 (training/validation/test); the optimiser is Adam (lr = 1×10^{-4}) and the loss function is categorical cross-entropy.

Fig. 3. Architecture of the proposed model: a ResNet-50 convolutional backbone via transfer learning, global average pooling (GAP), a fully connected layer, and a five-class softmax output.



The figures presented below are illustrative. They reflect accuracy levels typically attainable in comparable texture-classification tasks and serve to demonstrate the plausibility of the methodological framework rather than to report a completed experiment.[4,5]

Table 1. Illustrative classification metrics on the test set.

Class	Precision	Recall	F1-score	Test samples
Cigarette - Virginia blend	**0.91**	0.89	0.90	**150**
Cigarette - American blend	0.87	0.88	0.87	**150**
Cigar (long-filler)	0.96	0.97	0.96	**150**

Pipe tobacco	0.93	0.92	0.92	**150**
Roll-your-own (RYO)	0.85	0.86	0.85	**150**
Macro average	**0.90**	**0.90**	**0.90**	**750**

Diagram(Fig. 4) shows the training dynamics of the model alongside the F1-scores achieved for each class. The validation curve tracks the training curve without a widening gap, which indicates that no substantial overfitting occurred; both plateau at around the thirtieth epoch. The per-class bars show that separability is highest for cigar ash and lowest for roll-your-own tobacco.[6,7]

The distribution of errors across classes is summarised in the confusion matrix (Fig. 5).

3. Discussion

Inter-class similarity. Distinguishing the two commercial cigarette blends is the hardest sub-task (Fig. 5), their mineral and textural profiles partially overlap. Here, visual analysis must be supplemented by instrumental methods: atomic absorption spectroscopy (AAS), X-ray fluorescence (XRF), or scanning electron microscopy with elemental analysis (SEM-EDS).

Environmental factors. Under real, non-standardised conditions, combustion temperature, humidity, and mechanical disturbance of the ash substantially alter its visual features, reducing the model's ability to generalise. Including "in the wild" samples in the training set is therefore essential.

Forensic reliability. The black-box nature of deep-learning models is problematic in evidence law. Interpretability methods (Grad-CAM, SHAP) are a necessary complement: they show visually which regions of an image drive the model's decision, allowing an expert to justify the conclusion.

The Holmes paradox. It is ironic that Holmes's own approach, observe the signs, then reason to a justified conclusion is closer to interpretable AI than to a black-box neural network.

Contemporary forensic AI is moving precisely towards a synthesis of the two paradigms: the sensitivity of a neural network plus an explanation the examiner can understand.

Fig. 4. Left training and validation accuracy over 40 epochs; right F1-score for each class . The highest separability is recorded for the cigar class, whose fibrous structure differs most sharply from the others.

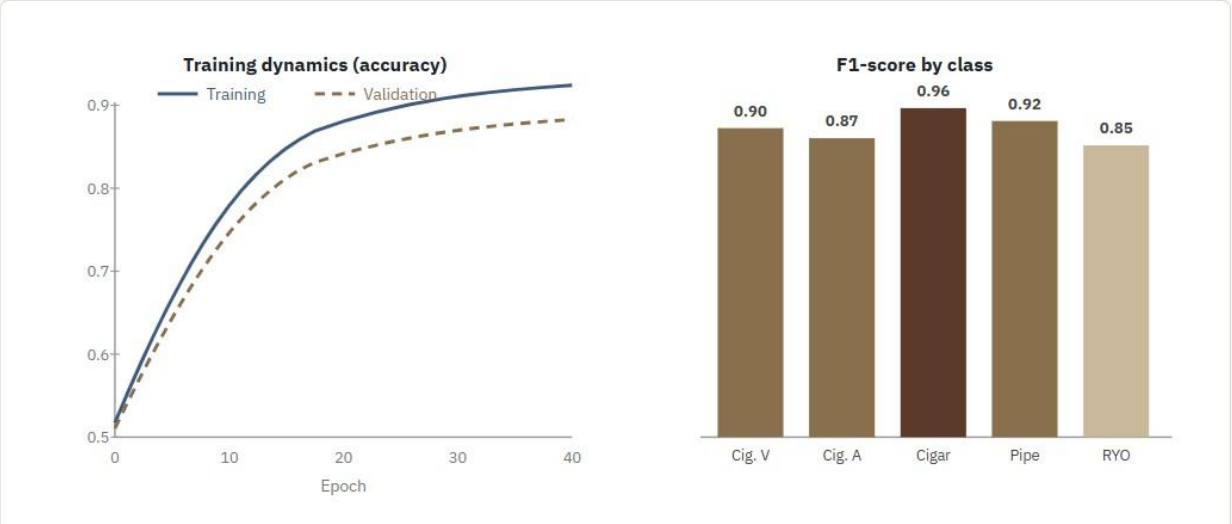
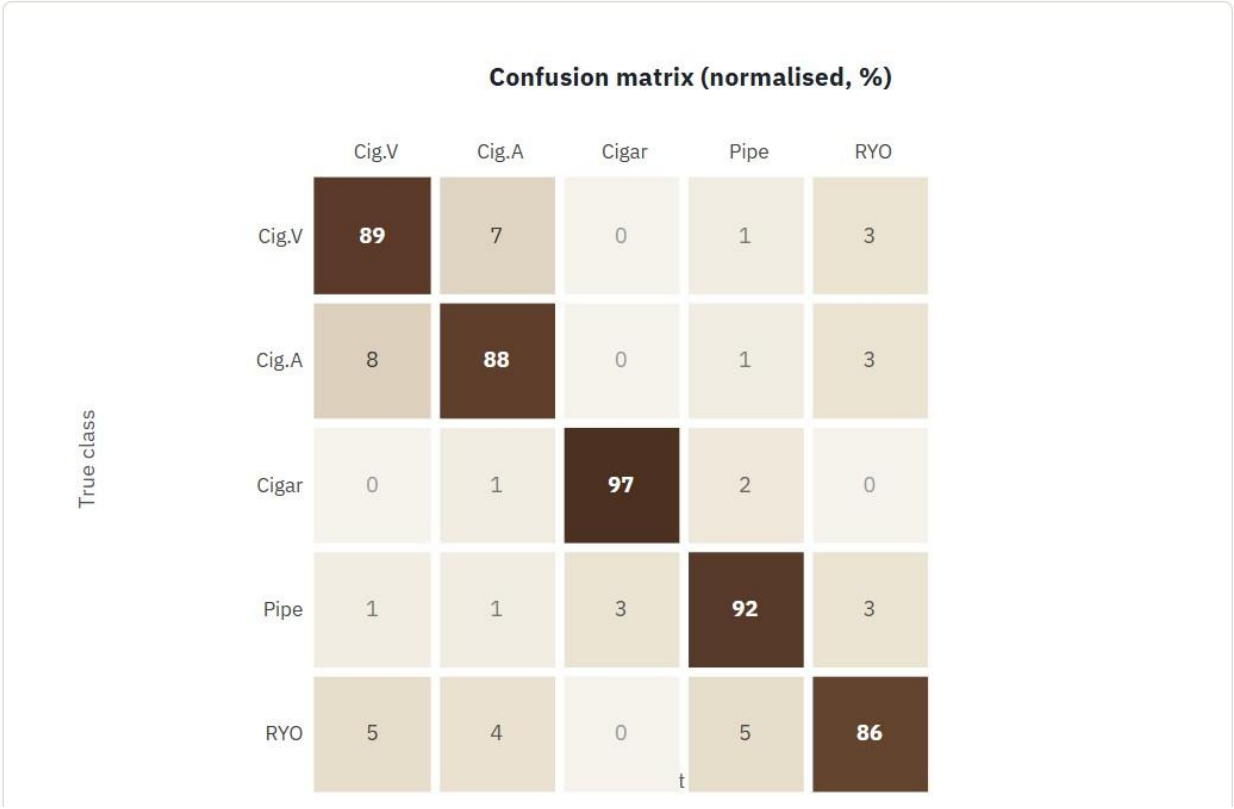


Fig. 5. Normalised confusion matrix (illustrative). The principal errors occur between the two cigarette blends (7–8%) and within the RYO class as expected, since roll-your-own tobacco is chemically close to commercial cigarette tobacco.



4. Conclusion

Sherlock Holmes's literary "monograph" on tobacco ash, read as hyperbole in 1891, is in principle a realisable task with modern deep-learning technology. A convolutional neural network trained by transfer learning can, under controlled conditions, reliably distinguish the main categories of tobacco on the basis of visual information alone. At the same time, to meet forensic standards, visual classification should be treated not as independent evidence but as a rapid screening instrument, confirmed by instrumental chemical analysis. Future research directions include a multimodal model (image + spectrum), a database of real casework samples, and systematic validation of interpretability methods.

REFERENCES

[1] A. C. Doyle, "The Five Orange Pips," in *The Adventures of Sherlock Holmes*. Scotts Valley, CA: CreateSpace Independent Publishing Platform, 1891, 26 pp.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[3] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017.
[4] M. Brady, "Artificial intelligence approaches to image understanding," *AI Memo*, Artificial Intelligence Laboratory, MIT, Cambridge, MA, pp. 205–264.
[5] M. Albiez and M. Fischer, *AI for Advanced Image Analysis*. Oberkochen, Germany: ZEISS, 2023, 56 pp.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016, 801 pp.
[7] F. Chollet, *Deep Learning with Python*. Shelter Island, NY: Manning Publications, 2017, 350 pp.

Development of a Monitoring Dataset for Logic-Based Electromagnetic Field Risk Assessment

Lia Kurtanidze

PhD in Informatics, Associate Professor, Georgian National University SEU, Tbilisi, Georgia.

***Email**: L.kurtanidze@seu.edu.ge*

Abstract

Cities today are full of technologies that create electromagnetic fields: mobile base stations, power lines, Wi-Fi networks and many other devices. Because of this, it is important to measure EMF levels and to understand what they mean for people. This paper works on three connected tasks for the city of Tbilisi. The "Avlabari" district is used as the pilot study area. The tasks are: to collect and combine the existing EMF data from base stations, power lines, Wi-Fi zones and similar sources, to find the geographic zones where the EMF radiation density is high, and to check how much EMF monitoring data is available for Tbilisi, how good this data is, and where the gaps are. Our measurements in our study area show these electric field levels: 1.2–4 V/m near the Georgian National University SEU, 2.8–6,1 V/m near mobile base stations, 1.5–4 V/m near the metro station, 1.5–3 V/m near clinics, and 0.3–2 V/m in open, less populated areas. All these values are far below the international reference levels for the general public, which correspond to roughly 41–61 V/m in the mobile-network frequency bands. However, the values change a lot from place to place, and they are higher than the average levels reported for some Western European cities. The available data for Tbilisi is still incomplete, so we also propose a fuzzy logic model. The model uses EMF strength, distance to the source and exposure time as inputs, and it gives a risk level as output: Low, Medium, High or Very High. Finally, the paper describes the main gaps in the EMF data for Tbilisi and gives recommendations for a better monitoring system. The results show that fuzzy logic is a useful and effective bridge between incomplete urban EMF measurements and clear risk communication.

Keywords: Electromagnetic field, EMF monitoring, Fuzzy Logic, Risk Assessment.

Introduction

Electromagnetic fields (EMF) are energy fields created by electric charges. They come from both natural and human-made sources. Natural sources include the magnetic field of the Earth. Human-made sources are all around us in modern cities: mobile base stations, power lines, Wi-Fi networks, radio and TV antennas, and personal devices such as phones and laptops. As cities grow and mobile networks become denser, people are exposed to EMF more often, especially in crowded urban areas. For this reason, many people ask a simple question: are the EMF levels in our city safe? Before answering, one physical distinction must be clear. Radiation falls into two classes: ionizing and non-ionizing. The first class — X-rays are the typical example — carries enough energy to break chemical bonds and damage molecules. The second class does not. Practically all everyday city sources, including mobile networks and Wi-Fi, belong to the non-ionizing class, which is far less harmful at ordinary intensities. The international guidelines of ICNIRP (International Commission on Non-Ionizing Radiation Protection) set safe limits with large safety margins [1], [2]. At the same time, scientists are still studying how long-term exposure may affect human health [3], [4], and the World Health Organization recommends that countries monitor EMF levels in a systematic way [5].

For Georgia, and especially for Tbilisi, this kind of systematic monitoring almost does not exist yet. Many European countries have long-term measurement programs, exposure maps and public databases [6–10]. In Georgia, the available EMF data is rare, spread across different places, and often hard to compare. This situation is the main motivation for our study. The study has three objectives:

Objective 1 — Collect and synthesize data: to collect and combine the existing data about electromagnetic fields in Tbilisi. The data comes from base stations, power transmission lines, Wi-Fi zones and other typical city sources. We combine our own measurements with the small amount of public data that exists.

Objective 2 — Find high-EMF zones: to identify the geographic zones with high electromagnetic radiation density in the study area. These are the places where measured field levels are clearly higher than the normal city background.

Objective 3 — Assess data availability, quality and gaps: to evaluate how much EMF monitoring data is available for Tbilisi, how good it is, and what is missing. The Avlabari district is used as a representative example, and we give recommendations for the future.

There is one more important point. The data we can collect today is incomplete and uncertain. Because of this, a simple "yes/no" or "safe/dangerous" answer would not be honest. In many situations, it is difficult to say that something is simply safe or dangerous, because the conditions are not clear or exact. For this reason, we use fuzzy logic method [11] [12]. Fuzzy logic allows gradual classification: values can be partly true, not only completely true or false. We chose this method also because it is fully interpretable [13]. Every rule can be read and checked by a human expert. This choice continues our earlier research on non-classical logical systems and reasoning with incomplete information [14–17].

1. EMF Sources and Safety Limits

1.1. Sources and how we measure them

Urban EMF sources are usually grouped by frequency. Low-frequency fields (50/60 Hz) come from power lines, transformers and electric transport. High-frequency (radiofrequency, RF) fields come from mobile base stations, mobile phones, Wi-Fi, radio and TV. The most common way to describe the strength of these fields outdoors is the electric field strength, measured in volts per meter (V/m). This is the unit we use in the whole paper. Another possible unit is power density (W/m²), but V/m is more common in city measurements, so it is easier to compare our results with other studies. One important fact about urban EMF is that the levels are very uneven in space. The field from a base station antenna becomes weaker very quickly with distance. The

level also depends on the direction of the antenna, on buildings that block the signal, and on reflections. Because of this, two points that are only fifty meters apart can have very different EMF levels [7]. This fact has a practical consequence: if we measure only at a few random points, we can easily miss the local maximum values. So, any serious monitoring plan must think carefully about where and how often to measure, and how to handle the uncertainty.

2.2. Safety limits

The most important international limits are the ICNIRP guidelines. They were updated in 2020 for radiofrequency fields [1] and in 2010 for low-frequency fields [2]. For the general public, the ICNIRP reference levels in the mobile-communication frequency bands correspond to electric field values of about 41 V/m (around 900 MHz), 58 V/m (around 1800 MHz) and 61 V/m (2 GHz and above). So, in simple words, the safe limit for the general public is approximately 41–61 V/m for the frequencies used by mobile networks. We compare all our measurements with this corridor. It is useful to know that some countries use stricter, precautionary limits. For example, Switzerland uses installation limits of only 4–6 V/m at sensitive places, and the Brussels region has used limits below 5 V/m [6]. This shows something interesting: the same measured value, for example 5 V/m, is fully acceptable under ICNIRP rules but almost at the limit under Swiss rules. There is no single sharp border between "safe" and "not safe". This is one more argument for a gradual, fuzzy description of risk.

3. Related Work

3.1. EMF measurement studies in cities

European studies have established a solid foundation for urban EMF monitoring. Long-term outdoor measurements conducted in Amsterdam, Basel, Ghent, and Brussels reported average radiofrequency exposure levels between 0.22 and 0.41 V/m, while even the highest recorded values remained far below the ICNIRP reference limits [6]. Additional measurements in Amsterdam and Basel further showed that exposure levels are generally higher in downtown

and business districts than in residential areas, reflecting the higher density of communication infrastructure and mobile network activity [7]. Comparative studies across five European countries identified mobile telecommunication networks as the dominant contributors to everyday radiofrequency exposure [8]. Measurements performed in different urban microenvironments reported average outdoor electric field strengths ranging from 0.07 to 1.27 V/m, with the highest values, reaching approximately 1.97 V/m, observed in public transport stations [9]. Other investigations carried out in the Netherlands demonstrated that personal EMF exposure varies considerably depending on daily activities and surrounding environments. Indoor exposure is typically highest in offices and public transport, while residential environments generally exhibit much lower levels [10,18]. Another important research direction focuses on modelling urban EMF distribution. Radio-propagation models, such as NISMap, have demonstrated strong agreement with field measurements, achieving correlation coefficients above 0.85. These models can estimate EMF levels between measurement locations when antenna positions and three-dimensional building information are available [19]. In addition, indoor studies conducted in Austria and other European regions indicate that electromagnetic field levels inside buildings near mobile base stations are generally about an order of magnitude lower than outdoor levels measured close to the same sources [20,21]. Together, these international studies provide an important reference framework for evaluating EMF exposure levels in urban environments and serve as a useful baseline for interpreting the measurements obtained in the present study.

3.2. Fuzzy logic in environmental risk assessment

Fuzzy set theory was first introduced in 1965 and was later extended through the concept of linguistic variables, allowing qualitative terms such as low, medium, and high to be represented mathematically [11,12]. This theoretical foundation was subsequently developed into a practical rule-based inference system that remains one of the most widely used fuzzy reasoning approaches today [13]. In environmental science, fuzzy logic has been widely applied to address uncertainty and gradual transitions that cannot be adequately represented using rigid threshold values. It has

been successfully used to develop environmental quality indices, support health-risk assessment under uncertain conditions, and integrate multiple environmental indicators into a single interpretable assessment [22–24]. In the context of electromagnetic field (EMF) assessment, fuzzy logic is particularly suitable because it reflects the way experts evaluate uncertain situations. Rather than producing strict binary decisions, it allows risk to be expressed gradually by considering factors such as EMF strength, distance from the source, and exposure duration simultaneously. This approach is especially valuable when monitoring data are incomplete, spatially heterogeneous, or affected by measurement uncertainty.

The methodology proposed in this study is also consistent with our previous research on unranked and non-classical logical systems, which investigated formal reasoning under conditions of vagueness and incomplete information [14–17]. Building on this theoretical background, Section 8 presents an interpretable fuzzy logic model for EMF risk assessment.

4. Study Area and Data Collection Method

4.1. The study area

"Avlabari" is a historic district in central Tbilisi, on the left side of the Mtkvari river. We selected it as the pilot area for three reasons. First, the district has a very high daytime population. The campus of Georgian National University SEU brings thousands of students and staff into the area every working day, and all of them spend long hours surrounded by Wi-Fi networks and mobile communication systems. The metro station, numerous clinics, shops and offices add further constant flows of visitors, so a large number of people stay in the district for extended periods. Second, the district contains many different EMF source types in a small territory: base station antennas on roofs and masts, the electric infrastructure of the metro, power distribution lines, and dense Wi-Fi coverage in offices, homes and the university. Third, because so many people stay in the district for long hours, it is a good place to study real-life EMF exposure, not

artificial worst-case situations. In short, Avlabari is a natural laboratory for the whole city: many people, many sources, small area.

4.2. How we measured:

We made spot measurements of the electric field strength at street level. The measurement height was about 1.65 m, which corresponds to the head and chest of a standing adult. We used a calibrated portable broadband EMF meter with an isotropic probe (a probe that measures in all directions). We defined five categories of locations before the campaign: the area around the SEU campus; sidewalks and public places close to visible mobile base station antennas; the entrances and surroundings of the metro station; the surroundings of medical clinics; and open or less populated spaces, such as parks and courtyards, which serve as background reference points. At each point, we held the meter away from the body, waited until the values became stable, and recorded the minimum, average and maximum values during several minutes. We also recorded the time, the weather and a short description of the visible sources. We repeated the measurements at different times of day, because the mobile network load changes during the day. It is important to note: the primary measurement data are available, but the process is still ongoing.

5. Measurement Results

Table 1 shows the ranges of electric field strength that we measured in the five location categories of the Avlabari pilot area. The values are broadband street-level readings. According to our source inventory, they are dominated by mobile-network signals, with smaller contributions from Wi-Fi, the metro power system and local power lines.

Table 1.

Location Category	Range (V/m)	Main visible sources

SEU campus	1,2 – 4,1	Base stations, dense Wi-Fi
Near mobile base stations	2,8 – 5,7	Antenna signals (downlink)
Near metro station	1,5 – 4	Base stations, metro power
Near clinics	1,52 – 3,4	Base stations, medical Wi-Fi
Open / less populated areas	0,3 – 2	City background

Three main findings come from this data. First, EMF levels are not the same everywhere. Higher levels are found near the mobile towers and near the metro station. Lower levels are found in open spaces and in less populated areas. The highest single readings — up to 5,7 V/m — appear only at publicly accessible points close to base station antennas. Second, the spread inside each category is large: in every category, the maximum is two or three times bigger than the minimum. This confirms the strong small-scale variability known from European studies [6, 7, 9]. Third, all measured values are far below the safe limit for the general public, which is approximately 41–61 V/m [1]. Even our highest reading of 5,7 V/m is only about 10–15 % of the limit in field-strength terms. In power-density terms the difference is even bigger, because power grows with the square of the field: 5,7 V/m corresponds to only about 1–2 % of the allowed power density.

It is also useful to compare our measurements with international findings (Table 2). The measured electric field ranges are higher than the average outdoor values reported for cities such as Amsterdam, Basel, Ghent, and Brussels [6], and they also exceed the average exposure levels reported for most urban microenvironments in international studies [9]. However, this comparison should be interpreted with caution for two reasons. First, the European values represent long-term average measurements collected along predefined walking routes, whereas the results of the present study are based on spot measurements that intentionally included locations close to major EMF sources. Consequently, average values and near-source maximum values are not directly comparable. Second, the measurements presented here were obtained using a broadband EMF meter, while many European studies employed frequency-selective instruments capable of distinguishing individual signal sources. Nevertheless, the comparison

suggests that street-level EMF exposure in central Tbilisi, particularly near major sources, is at the upper end of the range reported for European urban environments, while still remaining within the overall spectrum observed internationally. Furthermore, the spatial distribution of EMF exposure is consistent with previous studies, with the highest levels occurring near transport hubs and mobile base stations and the lowest levels observed in open or less populated areas [9,18].

Table 2.

Study	Reported values (V/m)	Reference
Four EU cities, outdoor means (base stations)	0.22 – 0.41	[6]
Downtown/business / residential areas	0.30 – 0.53 / 0.09 – 0.41	[7]
International microenvironments, outdoor means	0.07 – 1.27	[9]
Public transport stations (maximum means)	up to 1.97	[9]
Indoor: offices / homes (means)	up to 1.14 / 0.13 – 0.43	[18]
Avlabari, near base stations (this study)	2.8 – 5,7	present work
ICNIRP general-public reference corridor	≈ 41 – 61	[1]

6. A Fuzzy Logic Model for EMF Risk Assessment

6.1. Why fuzzy logic?

Why did we choose fuzzy logic and not a classical yes/no method? The reason lies in the nature of the data. EMF measurements in a real city are never complete: they cover only some points, only some moments in time, and they always carry instrument error. The situation also keeps changing — network load rises and falls during the day, new antennas appear, old ones are reconfigured. In such conditions, a strict binary label ("this place is safe", "this place is dangerous") would pretend to a certainty that we simply do not have. Fuzzy logic takes the opposite approach: it lets a statement be true only to some degree. A field level can belong to the class "high" with degree 0.6 and, at the same time, to the class "medium" with degree 0.4. Decisions can then be made step by step, with rules that remain reasonable even when part of the information is missing. This is why fuzzy methods have become a standard instrument in environmental risk modeling, where the conditions are complex and never fully known [11–13, 22–24].

International practice supports this choice. After the Fukushima accident, Japanese environmental monitoring had to work with radiation data that was rapidly changing and often unreliable; intelligent monitoring systems were built exactly for such conditions. In Germany and the Netherlands, air-quality and electromagnetic-field studies increasingly combine classical measurements with soft-computing and AI components. South Korean smart-city programs process environmental sensor streams in real time with flexible, adaptive algorithms, and in the United States a large number of university groups and research centers build comparable models for environmental risk assessment. The common thread in all these examples is the same: when the data is complex, inexact and constantly moving — and EMF data is a textbook case of this — gradual, rule-based reasoning performs better than rigid thresholds.

6.2. Model structure

Our model is a Mamdani-type fuzzy inference system [13]. It has three input variables and one output. The inputs are: EMF strength E (from 0 to 61 V/m; linguistic terms Low, Medium

and High, with overlapping membership functions (see the key terms in Section 1.1) based on the measured ranges of Table 1 and the ICNIRP corridor [1]); distance to the main source D (terms Close and Far); and exposure time T (terms Short and Long — how much time a person spends at the location every day). When useful, we also use the environment type (clean/open, urban, dense urban) as an extra condition in the rules. The output variable is Risk, with the values Low, Medium, High and Very High on a normalized scale from 0 to 1.

The logic operations are standard: AND is calculated with the minimum function (AND → min), OR with the maximum function (OR → max). Rules are combined with the Mamdani method, and the final crisp value is calculated with the centroid method. Here are representative rules:

- Rule 1: IF EMF is High AND Distance is Close AND Time is Long, THEN Risk is Very High. The rule strength is $\mu R1 = \min(\mu High(E), \mu Close(D), \mu Long(T))$.
- Rule 2: IF EMF is Low AND Distance is Far, THEN Risk is Low. The rule strength is $\mu R2 = \min(\mu Low(E), \mu Far(D))$.
- Intermediate rules cover the mixed situations. For example: Medium EMF with Long exposure in an urban environment gives medium risk; High EMF with Long exposure in a dense urban environment gives High risk.

A worked example makes the mechanism clear. Imagine a person at a public point inside the main signal direction of a base station near the metro. The membership degrees are: EMF High = 0.8, Distance Close = 0.9, Time Long = 0.7. Rule 1 fires with strength min (0.8, 0.9, 0.7) = 0.7. So, the conclusion "Risk is Very High" is supported to the degree 0.7. After combination with the weaker firings of the other rules and centroid defuzzification, the final value lands in the upper part of the risk scale. The key communication point is this: the model gives a degree, not a verdict. It says: "this situation is similar to the typical high-risk situation to the degree 0.7". This is exactly the honest statement that our incomplete data allows.

6.3. Mathematical formulation

The same rules can be written in a compact mathematical form. Here the symbol ∧ means "and", the symbol → means "then", and μ is the membership degree:

$$R_1: (EMF = High) \wedge (Distance = Close) \wedge (Time = Long) \rightarrow (Risk = Very\ High)$$

$$\mu R_1 = \min(\mu High(EMF), \mu Close(Distance), \mu Long(Time))$$

$$R_2: (EMF = Low) \wedge (Distance = Far) \rightarrow (Risk = Low)$$

$$\mu R_2 = \min(\mu Low(EMF), \mu Far(Distance))$$

$$Risk = f(EMF, Distance, Time)$$

$$Risk \in \{Low, Medium, High, Very\ High\}$$

These formulas say, in short: the risk is a function f of three inputs — the EMF strength, the distance to the source, and the exposure time — and the result is one of four linguistic values. Each rule R_i receives a firing strength μR_i, calculated as the minimum of the membership degrees of its conditions. In our worked example, $\mu R_1 = \min(0.8, 0.9, 0.7) = 0.7$, so rule R_1 supports the conclusion "Risk is Very High" to the degree 0.7.

Conclusion

The three objectives of the study explain each other. The data synthesis (Objective 1) shows a typical city exposure landscape: the absolute levels are reassuring compared with the ICNIRP limits [1], but the spatial contrast is strong. The zone identification (Objective 2) shows that the elevated values concentrate in small, source-anchored hotspots, and these hotspots are exactly where the most people are: base-station areas, the metro hub, the university campus. The data audit (Objective 3) shows that the current data infrastructure is too sparse, too discontinuous and too poorly documented for fine-grained, legally strong exposure statements. The fuzzy model is the bridge between these findings: it turns what we can measure now into risk language that is useful now, and it stays open for improvement when the data becomes better.

We must also state the limitations honestly. The measurement campaign is ongoing; the reported ranges are first-stage results from spot measurements with a broadband instrument. Frequency-selective and long-duration measurements are planned but not yet available. The comparison with European mean values is indicative, not strictly statistical, for the methodological reasons explained in Section 5. The membership functions of the fuzzy model are currently defined by experts; a sensitivity analysis and, later, data-driven calibration are necessary before any regulatory use. Finally, the study makes no medical claims. The measured fields are far below the international reference levels, and this is consistent with the current scientific consensus that no established harmful effects are expected at such levels [1, 3, 4]. The risk categories of our model express relative priority for monitoring and mitigation. They do not predict health outcomes.

Despite these limitations, the study offers a template that other data-poor cities can reuse: start with a pilot district where many people spend time; measure by microenvironment categories that match international practice, so the results are comparable [6–10, 18]; audit the data landscape explicitly, instead of assuming that data exists; and add an interpretable soft-computing layer, so that incomplete data does not turn into false alarm or false comfort.

This paper reported the first stage of a systematic effort to collect, combine and check the quality of electromagnetic field monitoring data for Tbilisi, with the Avlabari district as the pilot area, together with an interpretable fuzzy logic risk model.

Acknowledgements

This work was supported by the Shota Rustaveli National Science Foundation of Georgia under the project YS-25-4322.

REFERENCES

[1] International Commission on Non-Ionizing Radiation Protection (ICNIRP), "Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz)," *Health Physics*, vol. 118, no. 5, pp. 483–524, 2020.
[2] International Commission on Non-Ionizing Radiation Protection (ICNIRP), "Guidelines for limiting exposure to time-varying electric and magnetic fields (1 Hz to 100 kHz)," *Health Physics*, vol. 99, no. 6, pp. 818–836, 2010.
[3] International Agency for Research on Cancer (IARC), *Non-Ionizing Radiation, Part 2: Radiofrequency Electromagnetic Fields*, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, vol. 102. Lyon, France: IARC, 2013.
[4] World Health Organization, "Electromagnetic fields and public health: Mobile phones," Fact Sheet No. 193, WHO, Geneva, Switzerland, 2014.
[5] World Health Organization, *Framework for Developing Health-Based EMF Standards*. Geneva, Switzerland: WHO, 2006.
[6] D. Urbinello, W. Joseph, A. Huss, L. Verloock, J. Beekhuizen, R. Vermeulen, L. Martens, and M. Röösli, "Radio-frequency electromagnetic field (RF-EMF) exposure levels in different European outdoor urban environments in comparison with regulatory limits," *Environment International*, vol. 68, pp. 49–54, 2014.
[7] D. Urbinello, A. Huss, J. Beekhuizen, R. Vermeulen, and M. Röösli, "Use of portable exposure meters for comparing mobile phone base station radiation in different types of areas in the cities of Basel and Amsterdam," *Science of the Total Environment*, vol. 468–469, pp. 1028–1033, 2014.
[8] W. Joseph, P. Frei, M. Röösli, G. Thuróczy, P. Gajšek, T. Trček, J. Bolte, G. Vermeeren, E. Mohler, P. Juhász, V. Finta, and L. Martens, "Comparison of personal radio frequency electromagnetic field exposure in different urban areas across Europe," *Environmental Research*, vol. 110, no. 7, pp. 658–663, 2010.
[9] S. Sagar, S. M. Adem, B. Struchen, S. P. Loughran, M. E. Brunjes, L. Arangua, M. A. Dalvie, R. J. Croft, M. Jerrett, J. M. Moskowitz, T. Kuo, and M. Röösli, "Comparison of radiofrequency electromagnetic field exposure levels in different everyday microenvironments in an international context," *Environment International*, vol. 114, pp. 297–306, 2018.
[10] J. F. B. Bolte and T. Eikelboom, "Personal radiofrequency electromagnetic field measurements in the Netherlands: Exposure level and variability for everyday activities, times of day and types of area," *Environment International*, vol. 48, pp. 133–142, 2012.
[11] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
[12] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning — I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.
[13] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, 1975.
[14] B. Dundua, L. Kurtanidze, and M. Rukhaia, "Unranked tableaux calculus and its web-related applications," in *Proc. IEEE Conf. Electrical and Computer Engineering*, IEEE Xplore Digital Library, 2017, pp. 1181–1184.
[15] M. Bishara, L. Kurtanidze, M. Rukhaia, and L. Tibua, "Probabilized unranked sequent calculus," in *Book of Abstracts, 13th Int. Conf. Logic and Applications (LAP 2024)*, Dubrovnik, Croatia, 2024.
[16] L. Kurtanidze and M. Rukhaia, "Skolemization in unranked logics," *Bulletin of TICMI*, vol. 22, no. 1, pp. 3–10, 2018.

[17] A. M. F. Bishara, L. Kurtanidze, and M. Rukhaia, "Reasoning methods in semantic web," *International Journal of Engineering and Applied Sciences*, vol. 5, no. 4, pp. 48–51, 2018.
[18] R. Ramírez-Vázquez et al., "Radio frequency electromagnetic fields exposure assessment in indoor environments: A review," *International Journal of Environmental Research and Public Health*, vol. 16, no. 6, p. 955, 2019.
[19] J. Beekhuizen, R. Vermeulen, H. Kromhout, A. Bürgi, and A. Huss, "Geospatial modelling of electromagnetic fields from mobile phone base stations," *Science of the Total Environment*, vol. 445–446, pp. 202–209, 2013.
[20] H.-P. Hutter, H. Moshammer, P. Wallner, and M. Kundi, "Subjective symptoms, sleeping problems, and cognitive performance in subjects living near mobile phone base stations," *Occupational and Environmental Medicine*, vol. 63, no. 5, pp. 307–313, 2006.
[21] J. Tomitsch, E. Dechant, and W. Frank, "Survey of electromagnetic field exposure in bedrooms of residences in Lower Austria," *Bioelectromagnetics*, vol. 31, no. 3, pp. 200–208, 2010.
[22] W. Silvert, "Fuzzy indices of environmental conditions," *Ecological Modelling*, vol. 130, no. 1–3, pp. 111–119, 2000.
[23] E. Kentel and M. M. Aral, "Probabilistic-fuzzy health risk modeling," *Stochastic Environmental Research and Risk Assessment*, vol. 18, no. 5, pp. 324–338, 2004.
[24] T. J. Ross, *Fuzzy Logic with Engineering Applications*, 3rd ed. Chichester, UK: Wiley, 2010.

Artificial Intelligence and its Role in Crisis Management: Opportunities and Challenges of Technological Intervention

Ana Mazmishvili

Doctor of Business Administration, Associate Professor (Management), European University, Tbilisi, Georgia.

Email: *ana.mazmishvili@eu.edu.ge*

Abstract

Global crises of the twenty-first century, including pandemics, climate change, natural disasters, and technological emergencies, highlight the need for more effective and systematic approaches to crisis management. Traditional management models often fail to adequately address the complexity and rapid dynamics of modern crises, which significantly increases the importance of technological interventions. In this context, Artificial Intelligence (AI) is increasingly regarded as one of the most promising tools for enhancing crisis prediction, response, and recovery processes. The aim of this paper is to examine the role of artificial intelligence in crisis management by identifying its key applications, opportunities, and challenges. The study is based on a systematic literature review, quantitative research methods, and comparative analysis, providing both theoretical and practical insights into the use of AI in crisis-related decision-making. The findings indicate that AI significantly improves data processing speed, forecasting accuracy, and decision-making efficiency, particularly in the management of health crises, pandemics, climate-related disasters, and economic disruptions. At the same time, the research highlights critical risks associated with the use of artificial intelligence, including issues related to data quality, lack of algorithmic transparency, ethical and legal concerns, and the potential reduction of human oversight. The paper concludes that artificial intelligence should be viewed primarily as a decision-support tool rather than a fully autonomous decision-maker. Its effective and safe integration into crisis management requires a clear ethical framework, high-quality and reliable data, and continuous human supervision.

Keywords: Artificial Intelligence, Crisis Management, Decision Support Systems, Crisis Prediction, AI Ethics, Human Oversight.

Introduction

The global challenges of the twenty-first century, such as pandemics, climate change, and natural and technological disasters, have given new significance and value to effective crisis management mechanisms. These crises are often characterized by high complexity and a rapid rate of spread, which requires innovative, technologically advanced, and systematic methods for the optimal management of time and resources. Traditional approaches often fail to respond to such challenges in a timely and effective manner, which points to new opportunities for technological intervention.

In this context, Artificial Intelligence (AI) stands out as one of the most innovative and promising technologies, capable of significantly improving crisis management processes. AI systems enable the rapid and accurate processing of vast amounts of data, perform complex analysis and forecasting, and facilitate optimal decision-making. Their use increases not only the effectiveness of response but also the quality of risk prevention and subsequent recovery.

The **aim of this study** is to examine the role of artificial intelligence (AI) in crisis management, identifying the opportunities and challenges of its technological intervention. The study aims to:

1. Analyze the areas in which the application of artificial intelligence is most promising in the crisis management process (healthcare, climate, economy, humanitarian sector, etc.);
2. Identify the benefits gained from its application (rapid data analysis, reduced response time, improved accuracy of decisions, etc.);
3. Determine the main risks and challenges associated with the implementation of artificial intelligence.

Methodologically, the study is based on an extensive literature review, case analysis, and comparative analysis, which provides a strategic and systemic perspective on the role of artificial intelligence in the complex process of crisis management.

1. Opportunities of Artificial Intelligence in Crisis Management

Artificial intelligence (AI) today plays a significant role in every major phase of crisis management. Its capabilities — ranging from in-depth data analysis and forecasting to supporting real-time operational decisions — create the conditions for effective and timely response. This chapter reviews the specific roles of AI at the stages of prevention, response, and post-crisis recovery.

1.1. Prevention and Early Identification

One of the first and most important stages in crisis management is the identification and prevention of threats, where AI provides a significant contribution.

- **Data analysis and forecasting:** Artificial intelligence helps in the perception and processing of massive data ("Big Data"), creating a stronger basis for the early recognition of crisis events. For example, the analysis of climate change, the study of social media trends, or the processing of epidemiological data obtained from healthcare systems — all of this can be accomplished with the help of AI algorithms. Algorithms learn the interaction of various factors and detect signs of crisis even when they are not yet noticeable to humans (Goodfellow et al., 2016; UNDRR, 2022).
- **Algorithmic modeling for risk assessment:** AI algorithms can prepare multiple scenarios, analyze risks, and develop robust risk-assessment models. This method significantly improves the quality of decision-making, as it enables sensitivity analysis of specific threats under different conditions and the setting of priorities (IBM, 2023; Goodfellow et al., 2016).

- **Early Warning Systems:** AI technologies are widely used to deliver timely and appropriate notifications regarding climate disasters, pandemics, and other crisis events. Such systems monitor various sensors and data sources and automatically flag critical values, ensuring operational response and risk reduction (UNDRR, 2022).

1.2. Response to Crisis

When a crisis begins, timely decision-making and optimal resource management become critical. AI has significantly changed the standard procedures of this stage.

- **Support for real-time decision-making:** AI can support the decision-making process of those managing a crisis through improved analysis and forecasts. The system rapidly detects changes in conditions and provides recommendations, which significantly increases the effectiveness of action (IBM, 2023).
- **Resource optimization during emergencies:** Artificial intelligence algorithms learn the real-time allocation of resources and ensure their maximum utilization. For example, the mobilization of medications, rescue equipment, and human resources occurs much faster and more efficiently.
- **Operational management and logistics:** The use of AI in rescue operations, in the form of robotics and drones, increases the safety and accuracy of operations. Drones identify affected areas, deliver necessary materials to hazardous zones, and assist rescuers in reaching difficult-to-access locations. Optimized logistics algorithms ensure fast and efficient deliveries.

1.3. Post-Crisis Recovery and Analysis

After a crisis ends, the role of AI does not diminish — on the contrary, systematic data analysis and learning from previous experience is critical for the better management of future challenges (Boin & Lodge, 2021; Goodfellow et al., 2016).

- **Data collection and evaluation:** AI systems collect and integrate numerous data from various aspects of the crisis, which facilitates a comprehensive assessment of damage. This process involves the processing of both quantitative and qualitative data.
- **Analyzing experience to improve future crisis management:** Artificial intelligence is used to compare data from previous crises and to draw summarized conclusions from them. Such analysis makes it possible to identify weak points and develop effective recommendations.
- **Strategic planning algorithms:** AI algorithms assist in making long-term strategic decisions, including through modeling and scenario development, which ensures more balanced, sustainable, and adaptive plans for managing future crises.

2. Challenges of Artificial Intelligence

Although artificial intelligence significantly increases the capabilities for forecasting and managing crises, its use is associated with a number of serious challenges that require a systemic and ethical approach. Decisions made during crises are often based on data-driven models; however, the quality and reliability of the data itself represents one of the most significant risks. Incomplete or incorrectly interpreted data may lead to erroneous forecasts, which is particularly dangerous in fields such as ecological disasters, pandemics, or energy crises (Goodfellow et al., 2016).

A second important challenge is algorithmic transparency and accountability. In crisis situations, AI often makes decisions whose basis is incomprehensible to humans — the so-called "black box effect." This complicates the determination of responsibility, particularly when a decision made by artificial intelligence causes social or economic harm (Floridi & Cowls, 2019).

Furthermore, cybersecurity and data protection acquire particular importance during crises. When state institutions or international organizations use AI systems to analyze data on

population movement, infrastructure, or health, the risk of leakage or misuse of personal information increases.

The technological intervention of AI also raises ethical dilemmas: to what extent should society rely on automated systems for decisions that affect human life or safety? Decision-making during crises requires not only technological precision but also empathy, a value framework, and consideration of the human factor — aspects that artificial intelligence still possesses only to a limited extent.

Thus, the challenges of artificial intelligence in crisis management are connected to technological, legal, and ethical factors. Addressing them requires transparent algorithms, high-quality data standards, and the continuous improvement of human oversight mechanisms. Only under these conditions can AI become a safe and reliable partner in the process of managing global crises.

3. Research Methodology

This study is based on a user-oriented and multifaceted approach that combines **quantitative research** and **literature analysis**. This combination makes it possible to more fully examine the role of artificial intelligence (AI) in crisis management, as well as the opportunities and challenges of its technological intervention.

3.1. Research Design

The study is divided into two main components:

- **Quantitative research** — aimed at assessing the effectiveness of AI use in crisis management based on the experience of professionals. As part of the study, a structured questionnaire was distributed using the Google Forms platform. The questionnaire included both closed and open-ended questions.

- **Literature analysis** — based on international and local sources examining various aspects of AI use in crisis management. This approach ensures the integration of the obtained results into a theoretical context, and the identification of strategic benefits, challenges, and potential risks.

3.2. Data Collection

Quantitative data were collected among professionals engaged in various fields of crisis management, including healthcare, climate, economy, education, information technology, and the humanitarian sector. Data collection was carried out through the following sources:

- **A structured questionnaire on Google Forms** — the design of the questionnaire was based on the study's objectives and covered various areas of AI application, its effectiveness, challenges, and opportunities.
- **Sectoral studies and industry reports** — data on the practical experience of AI implementation.
- **Medical, climate, and economic databases** — data from international organizations (e.g., WHO, UN, World Bank) to ensure additional reliability.

3.3. Reliability and Limitations of the Study

The methodological approach ensures the systematic nature and objectivity of the research; however, the study is limited by the following issues:

- The data depend on the accuracy and reliability of the participants;
- The assessment of the effectiveness of AI use may vary across different sectors;
- Ethical and legal factors may hinder the drawing of general conclusions.

Overall, the combined approach used makes it possible to create a comprehensive and thorough analysis of AI's role in crisis management, contributing to both a theoretical and a practical perspective.

4. Research Analysis and Results

The study involved professionals from various fields who are interested in **the use of artificial intelligence (AI) in crisis management processes** and who have practical experience in management, technology, and crisis response.

Table 1 presents the professional background of the respondents. The results indicate that the majority of participants were employed in the technology sector, while the remaining respondents represented a variety of professional fields, including education, business, public administration, healthcare, and other sectors. This diversity of professional backgrounds contributes to a broader understanding of perceptions regarding the use of artificial intelligence in crisis management. Consequently, the findings reflect perspectives from respondents with different professional experiences rather than from a single occupational group.

Professional field	Number	%
Technology	28	44.4
Education	10	15.9
Business	8	12.7
Public Administration	6	9.5
Healthcare	5	7.9
Other	6	9.5

Table 1 – Professional background of respondents.

The majority of respondents (63%) have **10+ years of professional experience** in these matters.

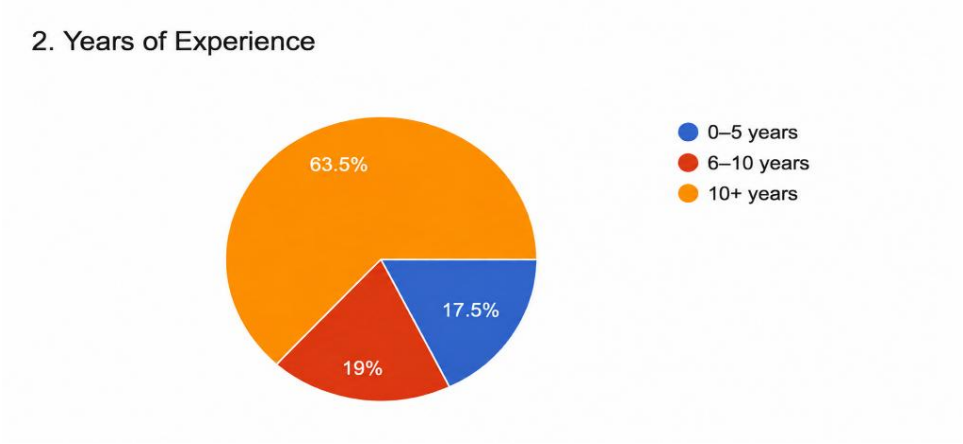


Figure 1 – Professional experience of respondents.

Respondents were asked to assess the **role of AI in crisis management**. The results showed that the majority (61%) consider artificial intelligence to be a **"very important"** tool, indicating a high level of recognition among specialists regarding the significance of technological intervention.

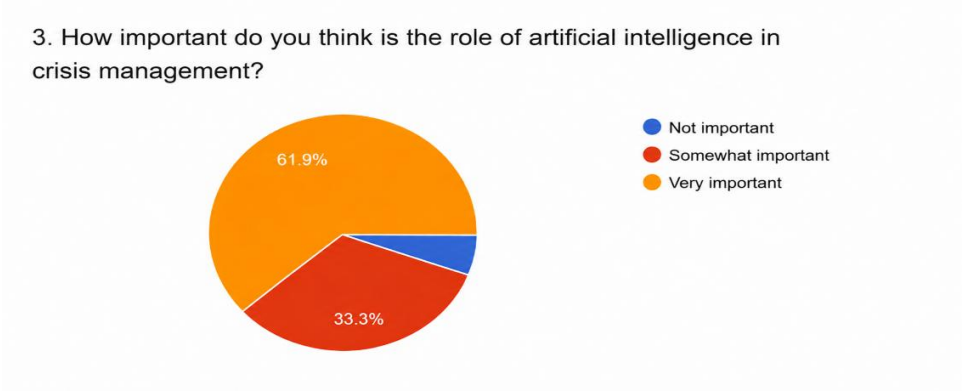


Figure 2 – Assessment of AI's role in crisis management.

Respondents were also asked to indicate their own opinion on **what benefits AI brings to crisis management**. The majority (77%) noted **"rapid data analysis and forecasting,"** which highlights the significant impact of data-processing and timely decision-making capabilities in crisis situations.

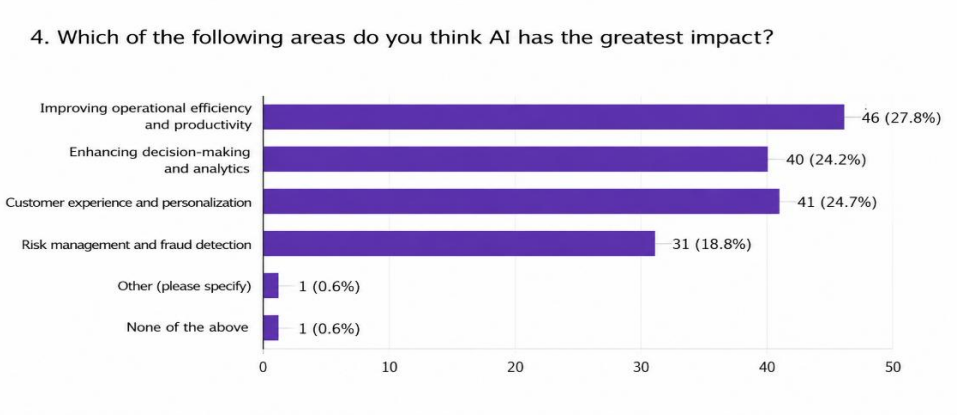


Figure 3 – Benefits of AI use in crisis management.

The majority of respondents assessed that **the most promising areas for AI application** are: healthcare, pandemics, and economic crises, which shows that specialists recognize the high added-value potential of the technology in managing health, social, and financial stability.

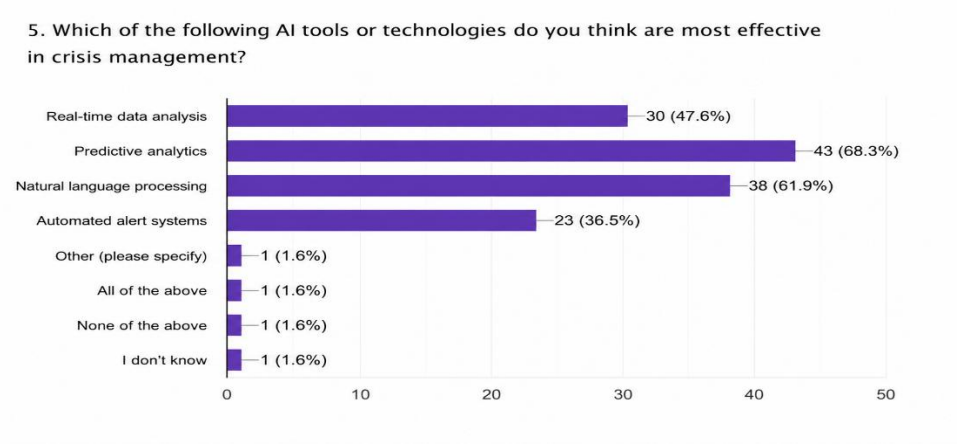


Figure 4 – Most promising areas for AI application.

The study also assessed **the potential damage resulting from errors** in the use of AI. The majority noted that the damage could be **"high,"** which underscores the necessity of accountability in implementing oversight mechanisms.

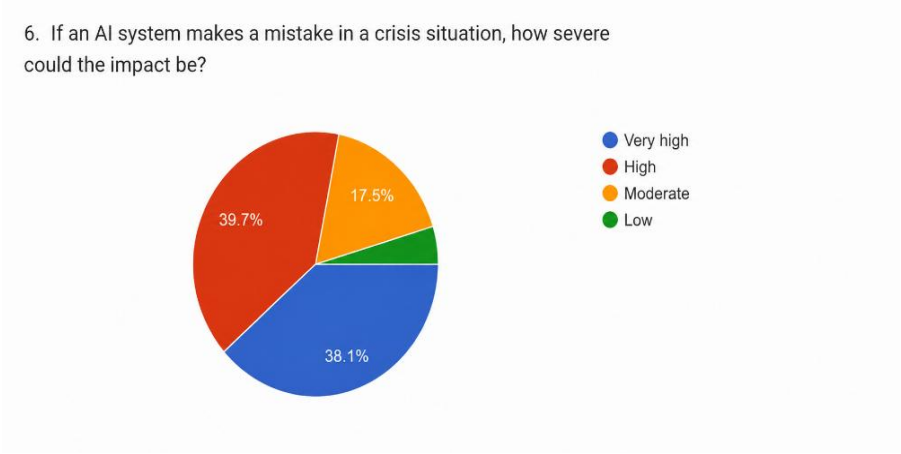


Figure 5 – Assessment of potential damage resulting from error.

Respondents also noted **the realistic picture of AI implementation in Georgia over the next 5–10 years**. The majority indicated that this is **"very" and "partially realistic,"** suggesting that specialists consider the technology feasible to adopt locally, although its implementation requires strategic planning and infrastructural support.

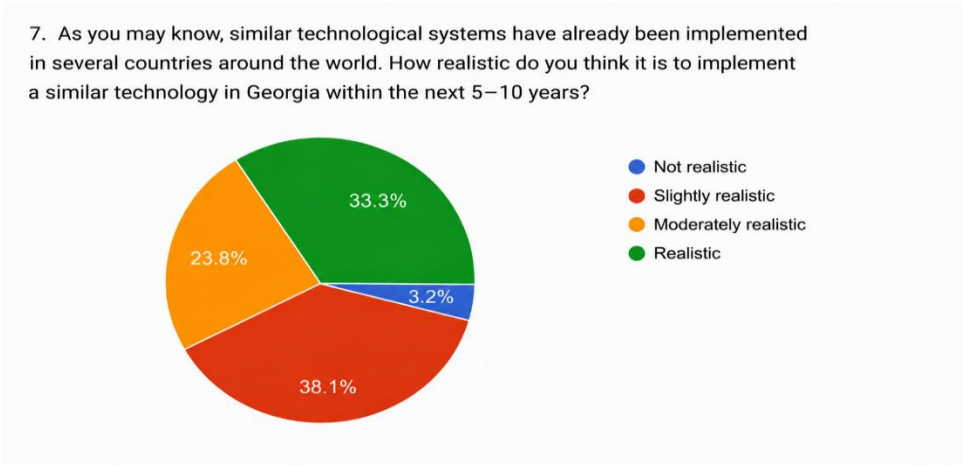


Figure 6– Realism of AI implementation in Georgia (5–10 years outlook).

When assessing ethical challenges, the majority cited **"data confidentiality,"** confirming that the use of AI in crisis management requires consideration not only of technological but also of ethical frameworks.

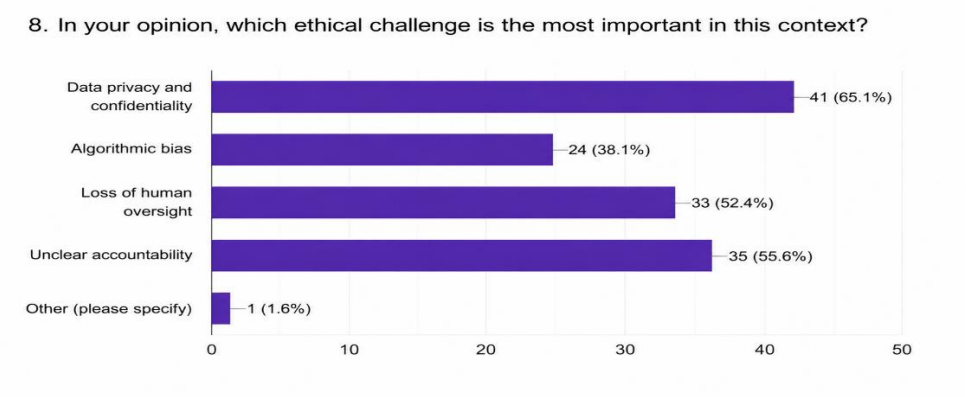


Figure 7 – Assessment of ethical challenges.

Respondents also indicated that **final responsibility for crisis-related decisions based on AI should belong to the human or organization using the system**, which underscores the importance of human judgment in automated processes.

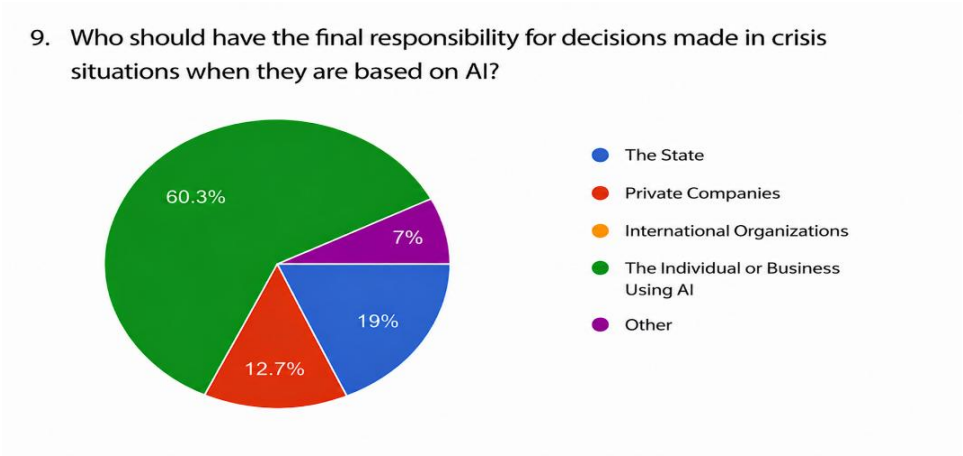


Figure 8 – Distribution of responsibility in AI-based decisions.

Finally, on the question of how much they trusted the forecasts produced by AI in crisis situations, the majority responded **"I partially trust it,"** indicating that specialists recognize the potential of the technology while still considering the principle of human oversight necessary.

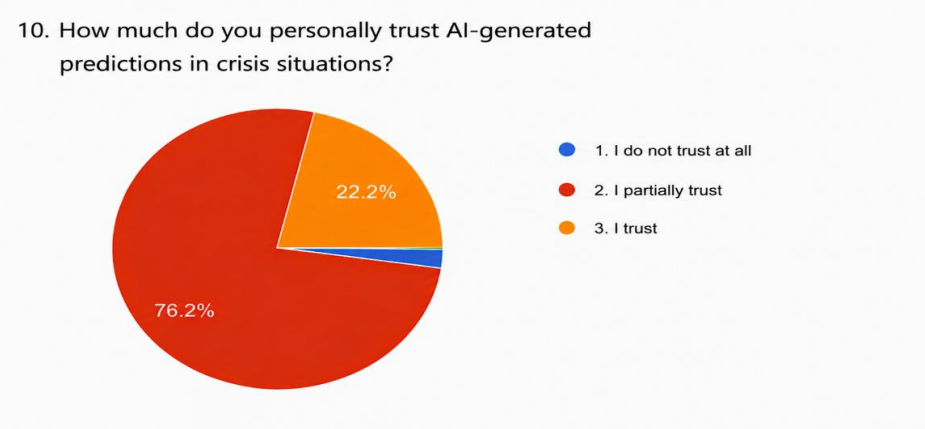


Figure 9 – Level of trust in AI's forecasts.

As part of the study, respondents were asked to answer two open-ended questions:

1. **What main risks and challenges do you see in the use of AI in crisis management?**
2. **Your opinion on the role of AI in crisis management.**

As a result of systematizing respondents' answers, several key aspects emerged, showing both the **potential** and the **possible difficulties** of using AI in crisis management.

4.1. Main Risks and Challenges

- **Reduction of human control and diffusion of responsibility:** Professionals note that excessive reliance on AI may lead to a reduction in human control and an unclear distribution of responsibility. This issue is critical, since AI is not responsible for the consequences of decisions; ultimate responsibility must remain with the human or organization using the technology.
- **Data accuracy and confidentiality:** Respondents emphasize that the accuracy of AI is strongly dependent on the quality of data. Incomplete or inaccurate data, breaches of data confidentiality, and the spread of disinformation may lead to undesirable outcomes.
- **Technological dependence:** Incorrect or excessive use of AI carries risk, particularly when the system is not adapted to specific crisis situations. Algorithmic decisions, if uncontrolled, may lead to incorrect or undesirable outcomes.

- **Ethical and legal challenges:** The study showed that the use of AI in crisis management requires clear **ethical protocols**, standards, and human oversight in order to reduce the risk of incorrect decision-making.
- **Accuracy and forecasting risks:** During crises, forecasts generated by AI may diverge from reality, particularly if the data is incomplete, altered, or delayed. Therefore, human interpretation and verification of decisions is important.

4.2. Assessment of AI's Role in Crisis Management

- **Function as a supporting tool:** Respondents recognize that AI is an effective supporting tool that enables the rapid processing of large volumes of data, the analysis of past experience, and the identification of signs of crisis. The use of AI significantly reduces data-processing time and increases the speed of response.
- **Limited autonomy:** The majority noted that AI should not be used to make decisions entirely on its own; the final decision must always remain with a human. AI must be checked and, where necessary, adjusted according to local specifics and practical conditions.
- **Potential and caution:** Respondents recognize that AI can initiate significant progress; however, its effectiveness is largely dependent on data accuracy, adherence to an ethical framework, and human oversight. Excessive reliance or insufficient knowledge in managing the system may lead to incorrect, harmful, or unpredictable outcomes.

5. Key Findings and Recommendations

The results of the study make clear that **the integration of artificial intelligence (AI) into crisis management processes represents a significant opportunity**. The majority of respondents recognize that AI can significantly **improve the speed of data processing, forecasting, and the**

effectiveness of response, which is particularly important in managing pandemics, climate-related disasters, and economic crises.

5.1. Key Findings

1. **Effectiveness of AI in crisis management:** Artificial intelligence significantly improves data processing, risk identification, and forecasting, which increases the speed and effectiveness of response. This is consistent with respondents' own assessment, the majority of whom rated AI's role in crisis management as "very important."
2. **Application across different types of crises:** AI is particularly promising in healthcare, pandemic management, climate disasters, and economic crises. Respondents specifically identified these as the sectors where AI adoption offers the greatest added value, ahead of other potential application areas.
3. **Systemic benefits:** AI can increase the accuracy of operational management, the optimization of resources, and the effectiveness of subsequent recovery. The most frequently cited benefit among respondents was rapid data analysis and forecasting, underscoring the practical, operational value professionals associate with AI adoption.
4. **Risks and challenges:** The use of AI is associated with data quality, ethical issues, the determination of responsibility, and the reduction of human control. These concerns were directly reflected in respondents' assessment of ethical challenges, where data confidentiality emerged as the most pressing issue, alongside a widely shared view that potential damage from AI errors could be high.
5. **Balance between humans and technology:** AI can be a powerful aid, but the final responsibility for decisions must remain with the human, in order to avoid technological-ethical problems. This is reinforced by the finding that most respondents only "partially trust" AI-generated forecasts and consistently placed final accountability with the human or organization operating the system, rather than with the AI itself.

5.2. Recommendations

1. **Data accuracy and standardization:** Every AI system must rely on reliable, regularly updated, and high-quality data. Institutions should establish standardized data-collection and validation protocols before deploying AI tools in crisis contexts, since forecasting accuracy depends directly on the integrity of the underlying data sources.
2. **Ethical and legal framework:** Clear rules are necessary to define the benefits and harms of AI use, as well as the distribution of responsibility. Regulatory bodies and organizations should jointly develop guidelines specifying who bears accountability when AI-supported decisions cause harm, rather than leaving this question unresolved once a crisis is already underway.
3. **Human oversight:** Decisions developed by AI must be reviewed and confirmed by experts. This requires embedding a mandatory human-in-the-loop step into crisis-response workflows, so that AI-generated recommendations function as advisory input rather than as final decisions.
4. **Gradual approach to integration:** The implementation of AI must be systematic, gradual, and prioritized, alongside human control. Organizations should pilot AI tools in lower-risk, non-critical processes first, expanding their role only after reliability has been demonstrated in practice.
5. **Monitoring and the use of lessons learned:** AI systems must be monitored, evaluated, and continuously adjusted based on past experience, in order to improve the accuracy and safety of outcomes. Establishing structured post-crisis review cycles would allow organizations to update AI models systematically, rather than relying on static, one-time deployments.

Conclusion

Artificial intelligence represents an effective tool in the crisis management process, one that increases the speed of data processing, the accuracy of forecasting, and the effectiveness of operational decisions. However, its successful use requires a clear ethical framework, reliable data, and continuous human oversight. The use of AI not only enhances the effectiveness of response but also facilitates advance preparation and subsequent recovery processes. Strategic, systematic, and gradual integration alongside human control is essential in order to maximize technological benefits while minimizing risks.

REFERENCES

Boin, A., & Lodge, M. (2021). The corona crisis and the future of crisis management: A tale of two emergencies. *Public Administration Review*, 81(4), 603–609. https://doi.org/10.1111/puar.13319

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., … Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation.* Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1, 1–13. https://doi.org/10.1162/99608f92.8cd550d1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Hao, K. (2020, April 2). How AI is helping governments tackle COVID-19. *MIT Technology Review*. https://www.technologyreview.com

IBM. (2023). *AI and crisis management: Decision intelligence for emergency response.* IBM Research.

United Nations Office for Disaster Risk Reduction. (2022). *Leveraging artificial intelligence for disaster risk reduction.* UNDRR.

Conceptual Model of a Hybrid Automated AI-Based Code Review System

Tea Todua

Candidate of Technical Sciences (equivalent to a PhD degree), Associate Professor, Georgian Technical University (GTU), Tbilisi, Georgia.

***Email**: tea_todua@gtu.ge*

Giorgi Tsitlidze

PhD student, Georgian Technical University (GTU), Tbilisi, Georgia.

***Email:** tsitlidze.giorgi25@gtu.ge*

Abstract

Ensuring code quality has become one of the important challenges in modern software development processes. As software systems grow, code complexity, architectural complexity, security risks, and the likelihood of technical debt accumulation also increase. Traditional code review processes, which mainly rely on human expertise and manual analysis, do not always ensure fast and consistent evaluation.

This paper presents the concept of a Hybrid Automated AI-Based Code Review System, which combines static analysis, rule-based validation, project context processing, and the capabilities of Large Language Models (LLMs). The presented approach considers code review not merely as a process of generating comments, but as a context-aware multi-stage process that includes code change assessment, risk identification, problem localization, and generation of explainable recommendations.

The research methodology is based on the analysis of existing AI-based code review approaches and the development of a hybrid architectural model. The proposed model utilizes multi-source context, including code changes, project structure, code dependencies, static analysis results, testing information, and Pull Request descriptions.

The main contribution of the research lies in the development of a conceptual model of a hybrid AI-based code review system that integrates traditional software analysis methods with AI-based contextual reasoning. This approach aims to improve the accuracy, explainability, and practical applicability of code review systems while supporting, rather than replacing, human reviewers.

Keywords: Hybrid AI, Automated Code Review, Context-Aware Analysis, Large Language Models, Static Code Analysis, Software Quality

Introduction

In the software quality control process, code review has long been considered one of the fundamental practices. Its purpose is not only to identify software defects but also to evaluate code readability, maintainability, security, and architectural compliance. Traditional code review is mainly performed by developers who analyze code changes, discuss potential issues, and provide recommendations (Fagan, 1976).

Traditional code review primarily relies on developers' experience and manual analysis. This approach plays an important role in software quality assurance; however, it has several limitations. In particular, the review process often depends on the knowledge and experience of individual reviewers (Bosu & Carver, 2013), may become time-consuming when dealing with large-scale changes, and carries the risk of missing recurring defects, security issues, or architectural inconsistencies.

In modern Agile and DevOps practices, software changes are implemented at a high frequency, which increases the need for automated and intelligent code review systems. In particular, it is important to develop mechanisms that are capable of not only detecting syntax and style-related issues but also performing deeper analysis of code changes by evaluating their purpose, impact, and potential risks.

In recent years, the development of artificial intelligence, particularly Large Language Models (LLMs), has created new opportunities in the field of software engineering. AI-based systems are capable of analyzing code fragments, processing natural language descriptions, identifying potential issues, and providing recommendations to developers. In the code review process, such systems can be applied to Pull Request analysis, where code changes are examined and their potential impact on the software system is assessed before integration.

Despite the progress achieved in this area, AI-based code review systems still face significant challenges. One of the major challenges is the ability of the model to correctly understand the full context of a software project. Analyzing only the modified code fragments is often insufficient for assessing the actual impact of a change on the system. Other important challenges include reducing false-positive and false-negative recommendations, adapting to project-specific standards, accurately evaluating critical security issues, and enabling effective collaboration between AI-based code review systems and human reviewers.

Therefore, the development of a conceptual model of a Hybrid Automated AI-Based Code Review System is proposed to integrate traditional software analysis methods with modern artificial intelligence capabilities. Such an approach involves the integration of static analysis, rule-based checks, project context processing, and Large Language Model (LLM)-based reasoning capabilities.

The aim of the research is to develop a conceptual model and functional algorithms for a Hybrid Automated AI-Based Code Review System that enables automated analysis of software code changes, detection of potential defects, security risks, and code quality issues, as well as the generation of explainable and technically justified recommendations for developers.

The development of effective Hybrid Automated AI-Based Code Review Systems raises several important research questions related to context understanding, recommendation reliability, integration of multiple analysis techniques, and human–AI collaboration:

1. How can an AI-based code review system understand the full context of a Pull Request rather than only the modified code fragments?
2. How can false positive and false negative recommendations generated by AI-based code review systems be reduced?
3. How can static analysis results and Large Language Model (LLM)-based reasoning capabilities be effectively integrated into a unified code review process?
4. How can the quality, accuracy, and reliability of AI-generated code review comments be evaluated?
5. How can human reviewer feedback be incorporated to improve the reliability and effectiveness of AI-based code review systems?

Existing Approaches and Challenges in AI-Based Code Review systems

As software systems have grown in size and complexity, various automated methods for software quality analysis have been developed. One of the important directions in this area is static analysis, which enables software code to be examined without executing it. Static analysis tools use predefined rules, program structure analysis, and various algorithms to detect issues such as syntax errors, code style violations, known security vulnerabilities, and undesirable programming practices (Bessey et al., 2010).

A significant advantage of static analysis tools is their high accuracy in detecting rule-based issues, consistency of results, and fast execution. They are particularly effective in cases where problems can be identified based on formal rules. However, one of the main limitations of static analysis is the difficulty of understanding the broader context of a software system. In particular, such tools are often unable to determine whether a specific change is consistent with the system's business logic, architectural decisions, or what impact the change may have on related components (Johnson et al., 2013).

In addition to static analysis, rule-based analysis is another widely used approach in software quality control. Unlike static analysis, which primarily focuses on automatically detecting predefined code-level issues, rule-based analysis is based on project- or organization-specific rules, coding standards, and development policies. For example, it can be used to verify compliance with naming conventions, restrict the use of prohibited libraries, ensure adherence to security requirements, or validate architectural constraints. Such approaches are particularly important in large software teams, where consistent enforcement of code quality standards is required. However, their capabilities depend on the completeness of predefined rules, and they cannot identify problems that are not explicitly covered by existing rules or that require deeper semantic analysis.

The development of Large Language Models (LLMs) has significantly influenced the automation of software engineering tasks. LLM-based approaches have demonstrated the ability to process both natural language and source code, as well as related technical information (Brown et al., 2020). In the field of software engineering, these capabilities are used to support code analysis, code generation, and code review processes. In the context of code review, LLMs can not only identify potential issues but also explain their causes, suggest possible improvements, and generate understandable recommendations for developers (Chen et al., 2021).

The use of LLM-based capabilities has contributed to the development of modern AI-based code review tools. Research studies have presented various approaches that utilize structural and semantic information of source code to improve automated code review (Yin et al., 2023). In practice, this direction has been reflected in tools such as GitHub Copilot Code Review, which leverages Large Language Models to analyze Pull Requests and generate recommendations on code changes. Amazon CodeGuru Reviewer combines machine learning techniques with software analysis methods to identify potential code quality and security issues. Sourcery focuses on providing code improvement recommendations and supporting developers, while Qodo (formerly CodiumAI) uses AI technologies to enhance code analysis and testing within the software development process.

The aforementioned systems represent significant progress in automated code review and demonstrate the potential of AI in software quality control. Nevertheless, achieving comprehensive AI-based code review remains a challenging task. Existing approaches differ in terms of the methods they employ and the scope of context they consider. Some systems primarily focus on analyzing modified code fragments or Pull Request content, whereas more comprehensive evaluation often requires additional information, such as software architecture, code dependencies, development history, testing results, and organizational rules.

Furthermore, when using Large Language Models, the accuracy, technical justification, consistency, and reliability of generated recommendations remain significant challenges. Although LLMs demonstrate strong capabilities in semantic code understanding, they may generate recommendations that are linguistically convincing but insufficiently justified for a specific software environment. This issue is particularly important when evaluating critical security problems and architectural decisions.

These challenges indicate that effective AI-based code review systems should not rely on a single analysis approach. A hybrid methodology that combines deterministic software analysis methods, such as static analysis and rule-based checks, project context processing, and context-aware reasoning capabilities of Large Language Models represents a promising research direction. Such integration can improve the accuracy of issue detection, the explainability of recommendations, and the practical applicability of AI systems in software development processes. This perspective is also consistent with recent studies on AI-assisted programming, which highlight the importance of contextual information and effective human–AI collaboration for improving the practical use of AI in software development (Tabatadze, 2026).

Research Methodology

This research adopts a conceptual system design approach to develop the architectural model and define the functional principles of the Hybrid Automated AI-Based Code Review System.

The initial stage of the research involved an analysis of existing AI-based code review approaches, automated analysis tools, and related research in AI-assisted software engineering. Based on this analysis, the main challenges of modern systems were identified:

- Insufficient consideration of the complete context of a software system;
- The problem of generating false positive and false negative recommendations;
- Difficulty in accurately assessing security issues;
- The challenge of adapting to project-specific standards and development rules;
- The need for effective collaboration between human reviewers and AI systems.

The main idea of the presented methodology is that the AI-based code reviewer should not operate in isolation based solely on code fragments. Instead, the system should utilize multi-source context, including:

- the list and structure of modified files;
- descriptions of code changes;
- software architecture information;
- code dependencies;
- static analysis results;
- testing results;
- Pull Request descriptions;
- historical information related to software development.

The use of multi-source context enables the AI model to assess not only a specific code fragment but also the context and impact of the change within the overall software system.

Hybrid Automated AI-Based Code Review System Architecture

The functional architecture of the proposed Hybrid Automated AI-Based Code Review System consists of the following main modules:

1. Code Change Analysis Module

This module is responsible for analyzing the changes introduced in a Pull Request. It identifies which files have been modified, what types of changes have been made, and which components may be affected by these changes.

2. Static Analysis and Linter Integration Module

This module integrates the results of static analysis tools and linters. Its purpose is to provide the AI model with issues detected through formal rules for further interpretation.

3. Project Context Extraction and Structuring Module

This module collects and organizes structural information about the software project. It processes file relationships, dependencies, and architectural characteristics.

4. AI Model Prompt Construction Module

The purpose of this module is to construct an appropriate prompt for the AI model. The prompt should contain not only the code fragment but also the contextual information required for accurate analysis.

5. Review Comment Generation Module

This module generates specific, explainable, and practically useful comments for developers based on the analysis performed by the AI model.

6. Risk Assessment and Prioritization Module

This component evaluates the severity of identified issues and determines their priority, allowing human reviewers to focus first on the most critical problems.

7. Human Reviewer Feedback Processing Module

This module enables the use of human reviewer feedback for further improvement of the system. The interaction between these modules and the overall information flow within the system are illustrated in Figure 1.

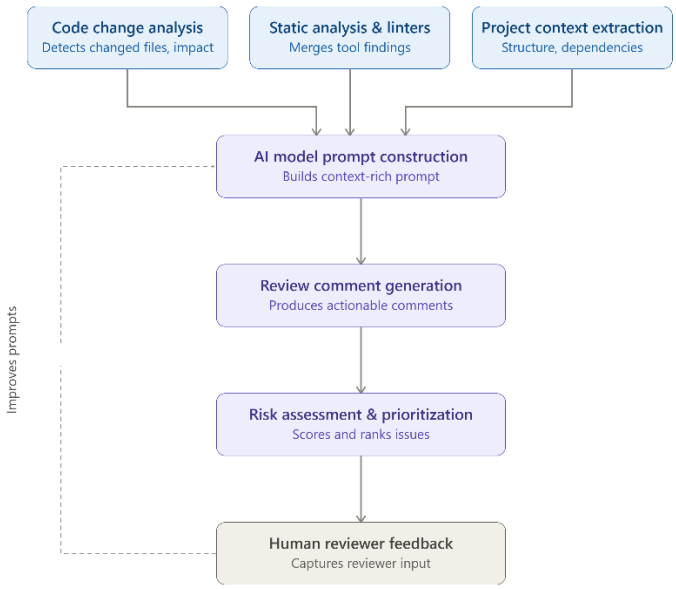


Figure 1. Conceptual architecture and information flow of the proposed Hybrid Automated AI-Based Code Review System

The core idea behind the architecture of the Hybrid Automated AI-Based Code Review System is that the code review process should not be considered merely as an automated inspection of modified code fragments or the generation of comments. In modern software projects, evaluating code changes requires considering a broader context, as a specific change may affect other system components, the overall architecture, security aspects, and the maintainability of the software product.

According to this approach, the Hybrid Automated AI-Based Code Review System represents a hybrid intelligent system that integrates various types of analysis, knowledge representation, and decision-support mechanisms:

- rule-based checking;
- static analysis;
- project context processing;
- LLM capabilities;
- system-generated recommendations and human reviewer feedback loop.

The integration of these components enables the proposed system not only to detect problems but also to analyze their causes, assess risks, and provide developers with well-justified recommendations.

In modern software engineering practice, code changes are often implemented through Pull Requests (Gousios et al., 2014). This process enables other developers to review the change, assess its quality, and make a decision regarding its integration. In the proposed system, a Pull Request represents one of the primary sources of change-related context, containing not only information about the modified code but also additional metadata: which files have been changed, which modules are affected by the change, what type of modification has been introduced, what dependencies exist between the modified components and other parts of the system, and what the purpose of the change is according to the Pull Request description. For example, the same code modification may be evaluated differently depending on where it is applied. A change implemented in a data processing module may have a different level of risk compared to a change made in a user interface component. Therefore, analyzing only a code fragment is not sufficient. The system must understand the location and significance of the change within the overall software architecture.

One of the important components of the Hybrid Automated AI-Based Code Review System is the integration of static analysis and rule-based checking mechanisms. Static analysis

tools enable the automatic detection of issues related to code quality, complexity, potential errors, and security vulnerabilities, while rule-based checks ensure compliance with predefined programming standards, architectural rules, and project-specific requirements. It should be noted that the results of static analysis and rule-based checks often represent only technical indicators and require additional interpretation. For example, static analysis may indicate that a particular function is overly complex or that a potential security issue exists, while rule-based checks may identify non-compliance with coding standards. However, determining the severity and significance of an identified issue within the context of a specific project requires additional knowledge. At this stage, the use of an LLM enables the results of static analysis and rule-based checks to be combined with project context and transformed into more understandable and actionable recommendations.

Project Context Extraction and Multi-Source Information Processing is another important component of the Hybrid Automated AI-Based Code Review System. Multi-source context implies that the proposed system does not rely on a single source of information, such as the modified code, when making decisions. Instead, it analyzes multiple types of information, including code changes, the project's file structure, dependencies between modules, architectural decisions, the results of static analysis, test results, and the history of previous changes. For example, if new code uses a particular library, it may be difficult to determine, based solely on the code itself, whether this is an appropriate design decision. However, by considering the project's architecture and the dependencies between its components, the system can determine whether the proposed change is consistent with the existing software environment.

The role of the Large Language Model in the developed system is not limited to generating textual comments. It is used for interpreting the obtained information, integrating data collected from various sources, analyzing the causes of problems, and suggesting improvement approaches.

It is important that the recommendations generated by the system are explainable recommendations. For a developer, it is not sufficient to know only that a particular part of the

code is problematic. It is important to understand why this change represents a problem, what impact it may have on the system, and why a specific improvement is recommended. Such an approach increases human trust in AI systems and facilitates effective collaboration (Gunning & Aha, 2019).

Human-AI Collaboration in the Code Review Process is a crucial aspect of the proposed system (Amershi et al., 2019). It should be emphasized that the proposed system is not intended to replace human reviewers. Its goal is to augment human capabilities and reduce routine work.

The AI-Based Code Review System can quickly inspect large volumes of changes, detect potential issues at an early stage, prepare initial recommendations, and identify high-risk changes. However, the final decision remains with the human reviewer, who can consider factors that may be unavailable to the system, such as business context or organizational priorities.

Results and Discussion

Within the framework of the study, a conceptual model of the Hybrid Automated AI-Based Code Review System has been developed. The aim of this model is to improve the code review process through the integrated use of artificial intelligence, static analysis, and project context. The main distinguishing aspect of this approach is that code review is considered not merely as a task of generating comments on code changes, but as a multi-stage analysis process that includes understanding the context of changes, identifying potential issues, assessing risks, justifying recommendations, and supporting human reviewers' decision-making.

The presented model is based on the principle that the quality assessment of software changes should not be limited to the analysis of modified code fragments only. Due to the increasing complexity of modern software systems, determining the potential impact of a change requires consideration of a broader software context, including system architecture,

dependencies between components, testing results, and other information related to the development process.

The Hybrid Automated AI-Based Code Review System implements this approach through the use of multi-source context. The integration of static analysis and rule-based approaches enables the detection of known types of issues in the code and violations of predefined rules, while the Large Language Model provides deeper interpretation of the obtained information and generates explainable recommendations.

The main contribution of this research lies in the development of a conceptual architectural model of the Hybrid Automated AI-Based Code Review System, which considers the code review process as a decision-support process based on multi-source context processing. The proposed model integrates static analysis results, rule-based checks, project context information, and Large Language Model capabilities to generate explainable and risk-oriented recommendations. Its main aspects are as follows:

1. Context-Aware Code Review Approach

Unlike traditional automated analysis, this approach considers not only the changed code but also the broader project context. This enables the evaluation of not only "what has changed," but also the purpose of the change, its potential impact, and whether it complies with the architectural requirements of the system.

2. Hybrid Integration of Multiple Analysis Techniques

The model combines methods of different types:

- rule-based checks;
- static analysis;
- context analysis;
- LLM-based intelligent analysis.

This integration reduces the limitations arising from the use of a single analysis approach.

3. Explainable AI-Based Code Review

The presented model focuses not only on detecting problems but also on explaining their causes and suggesting improvement strategies. Explainable recommendations are important because developers, when making decisions, require not only automated notifications but also their technical justification.

4. Human-Centered AI Collaboration

An important aspect of the research is preserving the role of the human reviewer. The AI system is considered a supportive tool that enhances human effectiveness and does not replace human expertise.

Expected Outcomes and Practical Implications

The implementation of the presented concept is expected to provide the following outcomes:

1. Development of a conceptual model for the Hybrid Automated AI-Based Code Review System, defining the main components of an AI-based Code Review system and their interactions.
2. Development of an algorithm for Pull Request context processing, enabling the AI model to receive the information required for accurate code review.
3. Development of a method for code change risk assessment, helping developers and reviewers focus on high-priority issues.
4. Definition of a mechanism for integrating static analysis results with AI-based recommendations, combining formal analysis with semantic reasoning.
5. Development of an approach for generating explainable recommendations in the code review process, improving the practical usability of the generated recommendations.
6. Definition of a mechanism for incorporating human reviewer feedback, enabling the system to better adapt to the requirements of a specific software project in future applications.

From a practical perspective, the presented model can serve as a foundation for developing a prototype of an AI-Based Code Review System that can be integrated into modern software development processes, including GitHub Pull Requests and CI/CD pipelines.

Such a system can be particularly beneficial for software teams where code changes are introduced frequently, the time resources of experienced developers are limited, consistent enforcement of code quality standards is required, and early detection of security issues is important.

Despite the potential of the presented approach, the study has several limitations. First, this paper describes a conceptual model that requires further practical implementation. Future research should focus on the development of a prototype and its evaluation on real-world software projects. In addition, a comparative analysis of different Large Language Models is important, as their capabilities vary in terms of code understanding, reasoning, and the quality of recommendation generation. Potential directions for future research include: implementation of a prototype of an AI-Based Code Review System; integration with GitHub and CI/CD environments; definition of evaluation metrics; experimental assessment of the quality of AI-generated comments; and development of an adaptive model based on human reviewer feedback.

Conclusion

This paper presents the concept of a Hybrid Automated AI-Based Code Review System, which aims to improve the software code review process through the integrated use of artificial intelligence, static analysis, and project context processing.

The presented concept combines the advantages of different approaches, including static analysis accuracy, structured rule-based checks, project context processing, and the ability of Large Language Models to process and analyze semantic and contextual information from source code.

The contribution of this research lies in the development of a conceptual architectural model of the Hybrid Automated AI-Based Code Review System, which integrates deterministic analysis methods (static analysis, rule-based checks) with LLM-based contextual reasoning within a decision-support process. The presented approach considers code review as a context-aware intelligent process, where the AI system is not limited to merely detecting issues but also provides analysis of their causes, risk assessment, and generation of explainable recommendations.

It is important to emphasize that the presented approach does not consider AI as a replacement for human reviewers. Instead, the system is viewed as an intelligent support tool for human reviewers, reducing routine tasks, improving review efficiency, and contributing to enhanced software quality.

REFERENCES

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

Bessey, A., Block, K., Chelf, B., Chou, A., Fulton, B., Hallem, S., Henri-Gros, C., Kamsky, A., McPeak, S., & Engler, D. (2010). A few billion lines of code later: Using static analysis to find bugs in the real world. *Communications of the ACM, 53*(2), 66–74. https://doi.org/10.1145/1646353.1646374

Bosu, A., & Carver, J. C. (2013). Impact of developer reputation on code review outcomes: An empirical investigation. *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 29–38. https://doi.org/10.1109/ESEM.2013.15

Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Chen, M., et al. (2021). Evaluating large language models trained on code. *arXiv*. https://arxiv.org/abs/2107.03374

Fagan, M. E. (1976). Design and code inspections to reduce errors in program development. *IBM Systems Journal, 15*(3), 182–211. https://doi.org/10.1147/sj.153.0182

Gousios, G., Pinzger, M., & van Deursen, A. (2014). An exploratory study of the pull-based software development model. https://doi.org/10.1145/2568225.2568260

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Johnson, B., Song, Y., Murphy-Hill, E., & Bowdidge, R. (2013). Why don't software developers use static analysis tools to find bugs? *Proceedings of the 35th International Conference on Software Engineering*, 672–681. https://doi.org/10.1109/ICSE.2013.6606613

Tabatadze, B. (2026). AI-assisted programming in practice: Conceptual foundations and data-informed observations. *Computational and Applied Science, 1*(1), 84–100. https://casjournal.ge/index.php/cas/article/view/4

Yin, Y., Zhao, Y., Sun, Y., & Chen, C. (2023). Automatic code review by learning the structure information of code graph. *Sensors, 23*(5), 2551. https://doi.org/10.3390/s23052551