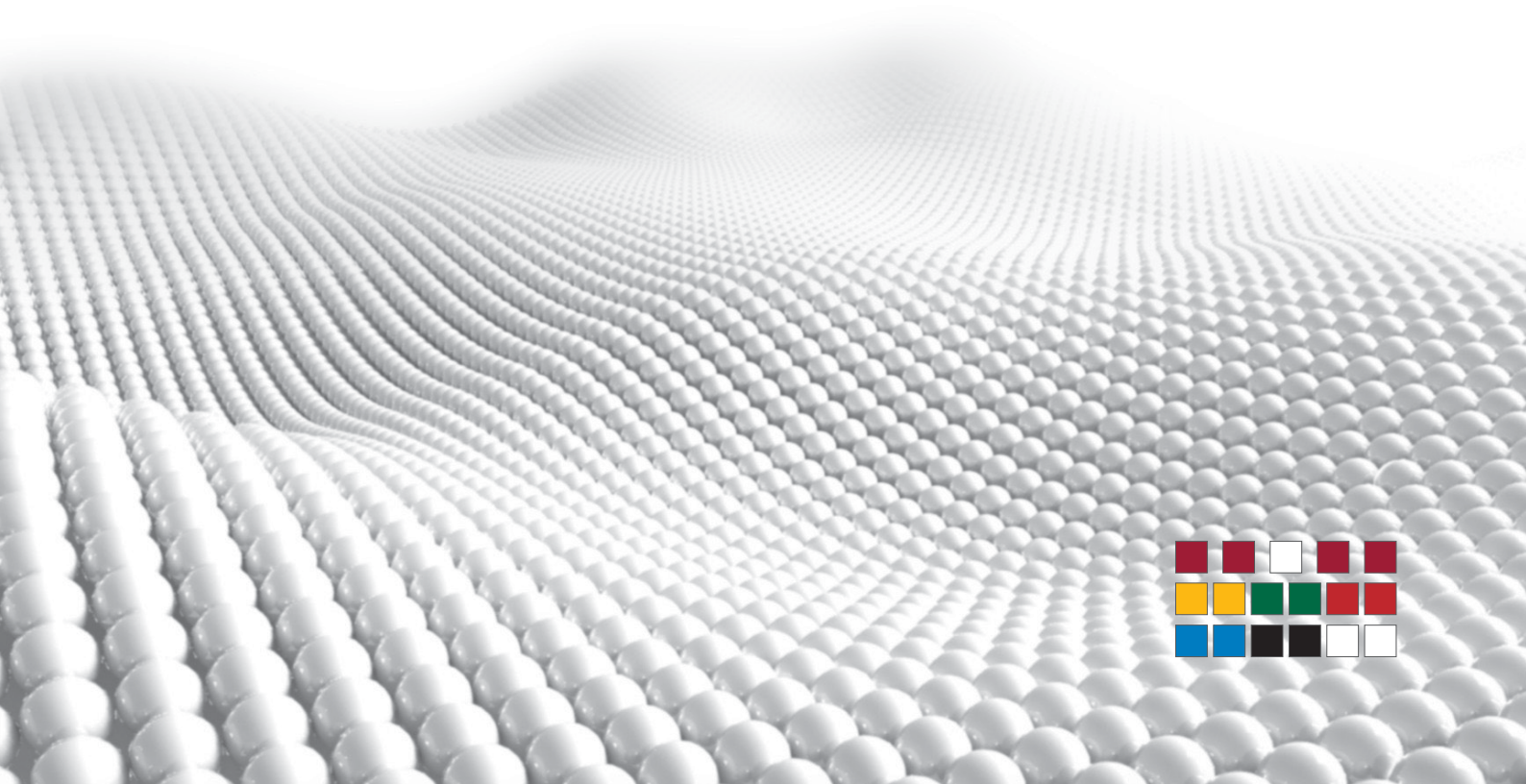
ISSN 2255-8950 (Online)
ISSN 2255-8942 (Print)

Volume 14 (2026)

No. 1

Baltic Journal of Modern Computing



CO-PUBLISHERS



Vilnius
University



UNIVERSITY OF
LATVIA



Latvia University
of Life Sciences
and Technologies



Institute of Mathematics and Computer Science
University of Latvia



VIDZEMES
AUGSTSKOLA

EDITORIAL BOARD

Co-Editors-in-Chief

Prof. Dr.habil.sc.comp. **Juris Borzovs**, Full Member of Latvian Academy of Sciences,
University of Latvia, Latvia
Prof. Dr. habil. **Gintautas Dzemyda**, Full Member of Lithuanian Academy of Sciences,
Vilnius University, Lithuania
Prof. Dr. **Raimundas Matulevičius**, University of Tartu, Estonia

Managing Co-Editors

Dr.sc.comp. **Jolita Bernatavičienė,** Vilnius University, Lithuania,
Dr.sc.comp. **Ēvalds Ikaunieks**, University of Latvia, Latvia
Prof. Dr. **Mubashar Iqbal**, Institute of Computer Science, University of Tartu, Estonia

Editorial Board Members (in alphabetical order)

Prof. Dr.sc.comp. **Andris Ambainis**, Full Member of Latvian Academy of Sciences,
University of Latvia, Latvia
Prof. Dr. **Irina Arhipova**, Latvia University of Life Sciences and Technologies, Latvia
Prof. Dr.sc.comp. **Guntis Arnicāns**, University of Latvia, Latvia
Assoc. Prof. Dr. **Mikhail Auguston**, Naval Postgraduate School, USA
Prof. Dr. **Liz Bacon**, University of Abertay, UK
Dr. **Rihards Balodis-Bolužs**, Institute of Mathematics and Computer Science,
University of Latvia, Latvia
Prof. Dr. **Eduardas Bareisa**, Kaunas University of Technology, Lithuania
Prof. Dr. **Romas Baronas**, Vilnius University, Lithuania
Prof. Dr.sc.comp. **Guntis Bārzdiņš**, Full Member of Latvian Academy of Sciences, University of Latvia
Prof. em. Dr.habil.sc.comp. **Jānis Visvaldis Bārzdiņš**, Full Member of Latvian Academy of Sciences,
Institute of Mathematics and Computer Science at University of Latvia, Latvia
Prof. Dr.sc.comp. **Jānis Bičevskis**, University of Latvia, Latvia
Prof. em. Dr.habil.sc.ing. **Ivars Biļinskis**, Full Member of Latvian Academy of Sciences,
Institute of Electronics and Computer Science, Latvia
Assoc. Prof. Dr. **Stefano Bonnini**, University of Ferrara, Italy
Dr.sc.comp. **Alvis Brāzma**, Foreign Member of Latvian Academy of Sciences,
European Molecular Biology Laboratory – European Bioinformatics Institute, UK
Prof. **Christine Choppy**, Université Paris 13, France
Prof. Dr.sc.comp. **Kārlis Čerāns**, Corresponding Member of Latvian Academy of Sciences,
Institute of Mathematics and Computer Science at University of Latvia, Latvia,
Prof. Dr. **Valentina Dagienė**, Vilnius University, Lithuania
Prof. Dr. **Robertas Damaševičius**, Kaunas University of Technology, Lithuania
Prof. Dr.sci. **Vitalij Denisov**, Klaipeda University, Lithuania
Prof. Dr.sci. **Kestutis Dučinskas**, Klaipeda University, Lithuania
Prof. Dr. **Ioan Dzitac**, Agora University of Oradea, Romania
Prof. Dr.habil. **Vladislav Fomin**, Vilnius University, Lithuania
Prof. **Sanford C. Goldberg**, Northwestern University, USA
Prof. Dr. sc. ing. **Jānis Grabis**, Riga Technical University, Latvia
Prof. Dr.habil.sc.ing. **Jānis Grundspeņķis**, Full Member of Latvian Academy of Sciences,
Riga Technical University, Latvia

Prof. Dr.habil. **Hele-Mai Haav**, Tallinn University of Technology, Estonia
Dr. **Nissim Harel**, Holon Institute of Technology, Israel
Dr. **Delene Heukelman**, Durban University of Technology, South Africa
Prof. em. Dr. **Kazuo Iwama**, Kyoto University, Japan
Prof. Dr.sc.comp. **Anita Jansone**, Liepāja Academy at Riga Technical University, Latvia
PhD **Oskars Java**, Vidzeme University of Applied Sciences, Latvia
Prof. Dr.habil.sc.ing. **Igor Kabashkin**, Corresponding Member of Latvian Academy of Sciences, Transport and Telecommunication Institute, Latvia
Prof. Dr. **Diana Kalibatienė**, Vilnius Gediminas Technical University, Lithuania
Prof. Dr.habil. sc.comp. **Audris Kalniņš**, Corresponding Member of Latvian Academy of Sciences, Institute of Mathematics and Computer Science at University of Latvia, Latvia
Assoc. Prof. Dr.phys. **Atis Kapenieks**, Riga Technical University
Prof. Dr. **Egidijus Kazanavičius**, Kaunas University of Technology, Lithuania
Prof. Dr.sc.phys. and math. **Vladimir Kotov**, Belarusian State University, Belarus
Adj. Prof. Dr. **Dmitry Korzun**, Petrozavodsk State University, Russia
Prof. Dr.habil. **Algimantas Krisciukaitis**, Lithuanian University of Health Sciences, Lithuania
Assoc. Prof. Dr. **Olga Kurasova,** Vilnius University, Lithuania
Prof. Dr. **Ivan Laktionov**, Dnipro University of Technology, Dnipro, Ukraine
Assoc. Prof. Dr, **Audronė Lupeikienė,** Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Vilnius University
Prof. Dr. **Raimundas Matulevičius**, University of Tartu, Estonia
Prof. Dr.habil.sc.ing. **Yuri Merkuryev**, Full Member of Latvian Academy of Sciences, Riga Technical University, Latvia
Prof. Dr.habil. **Jean Francis Michon**, retired from University of Rouen, France
Prof. Dr.habil. **Dalius Navakauskas**, Vilnius Gediminas Technical University, Lithuania
Prof. Dr.sc.comp. **Laila Niedrīte**, University of Latvia, Latvia
Assist. Prof. PhD **Anastasija Ņikiforova**, University of Tartu, Tartu, Estonia
Prof. Dr.sc.ing. **Oksana Nikiforova**, Riga Technical University, Latvia
Prof. Dr. **Vladimir A. Oleshchuk**, University of Agder, Norway
Prof. Dr.habil. **Jaan Penjam**, Tallinn University of Technology, Estonia
Assoc. Prof. PhD **Eduard Petlenkov**, Tallinn University of Technology, Estonia
Assoc. Prof. PhD **Ivan I. Piletski**, Belarussian State University of Informatics and Radioelectronics, Belarus
Prof. Dr.math. **Kārlis Podnieks**, University of Latvia, Latvia
Prof. Dr. **Boris Pozin**, Moscow State University of Economics, Statistics and Informatics (MESI), Russian Federation
Prof. Dr. **Tarmo Robal**, Tallinn University of Technology, Estonia
Prof. Dr. **Andreja Samčović**, University of Belgrade, Serbia
Prof. Dr.sc.eng. **Egils Stalidzāns**, University of Latvia, Latvia
Prof. Dr.sc.comp. **Leo Seļavo**, University of Latvia, Latvia
Prof. Dr. **Janis Stirna**, Stockholm University, Sweden
Prof. Dr.phil. **Jurģis Šķilters**, University of Latvia, Latvia
Prof. Dr.sc.eng. **Uldis Sukovskis**, Riga Technical University, Latvia
Prof. Dr.sc.comp. **Darja Šmite**, Blekinge Institute of Technology, Sweden
Prof. Dr. **Kuldar Taveter,** University of Tartu, Estonia
Prof. Dr.sc.comp. **Juris Vīksna**, Full Member of Latvian Academy of Sciences, Institute of Mathematics and Computer Science at University of Latvia, Latvia
Prof. Dr.sc.comp. **Māris Vītiņš**, University of Latvia
Prof. Dr.sc.ing. **Gatis Vītols**, Latvia University of Life Sciences and Technologies
Prof. Dr.art **Solvita Zariņa**, University of Latvia
Prof. em. Dr.rer.nat. habil. **Thomas Zeugmann**, Hokkaido University, Sapporo, Japan

Table of Content

Juan Peraza GARZÓN, Mónica Olivarría GONZÁLEZ, Yadira Quiñonez CARRILLO, Carmen Lizárraga BERNAL, Lucio Quirino RODRÍGUEZ, Alvaro Peraza GARZÓN

Immersive Virtual Environments in Higher Education: An Open-Source 3D World Adaptation Using Oculus Meta Quest 1–26

Asta KAZLAUSKIENĖ

Acoustic Features of Lithuanian Directives 27–49

Imelda ZADEJA

ADF^{2}S Feature Set of Anomaly Detection Framework Based on Knowledge Discovery 50–69

Palmi LAHE, Tarmo ROBAL

Digital Competencies and Barriers of Technology Adoption in Care Giving: A Systematic Literature Review 70–93

Daryna PYSHCHYK, Kuldar TAVETER, Mohamad GHARIB

An Approach for Engineering Software Ecosystems Based on an Industrial Case Study..... 94–120

Matīss ERIŅŠ, Zigurds MARKOVIČS

Analysis of Electroencephalogram Parameters to Determine Human Drowsiness..... 121–134

Arturs NIKULINS, Inese POĻAKA, Kaspars SUDARS

Evaluating Binary, Perlin Noise, and Physical Mark of Backdoor Attacks in Medical Image Classification..... 135–155

Cem ÖZKURT, Özkan CANAY, Ahmet KALA, Eyüp Altuğ TUNÇ, Nil Feyza ÖZDEMİR
Integrating Quantum Mechanics and Fuzzy Logic for Enhanced MCDM: A Case Study on Robot Evaluation..... 156–173

Maksims DIMITRIJEVS, Mārtiņš KĀLIS, Iļja REPKO
Exploring Hybrid Quantum-Classical Methods for Practical Time-Series Forecasting..... 174–189

Vasil CVETKOV, Christina MICKRENSKA
Uncertainty of a State Vertical Reference System below 1 mgpu – Is It Possible?..... 190–201

Oleksandr HALUSHKA, Viktor SHYNKARENKO
Neural Networks and Knowledge Integration: Ontology-Driven Text Encoding for CLIP..... 202–216

Aurelija KASPARAVIČIŪTĖ, Svetlana DANILENKO, Rūta SIMANAVIČIENĖ
Exploring the Connection Between Personality Dimensions and Learning Strategies in Mathematics..... 217–237

Vasil CVETKOV, Christina MICKRENSKA
Preliminary Reassessment Results of the Recent Vertical Movements of the Earth's Crust in Bulgaria..... 238–261

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 1–26
https://doi.org/10.22364/bjmc.2026.14.1.01

Immersive Virtual Environments in Higher Education: An Open-Source 3D World Adaptation Using Oculus Meta Quest

Juan Peraza GARZÓN, Mónica Olivarría GONZÁLEZ,
Yadira Quiñonez CARRILLO, Carmen Lizárraga BERNAL,
Lucio Quirino RODRÍGUEZ, Alvaro Peraza GARZÓN

Faculty of Informatics Mazatlan, Autonomous University of Sinaloa, Mazatlán Sinaloa, Mexico

jfperaza@uas.edu.mx, m.olivarria@uas.edu.mx, yadiraqui@uas.edu.mx,
carmen.lizarraga@uas.edu.mx, lquirino@uas.edu.mx, aperaza@uas.edu.mx

ORCID 0000-0003-3228-7037, ORCID 0000-0002-9335-5913, ORCID 0000-0002-7604-8532,
ORCID 0000-0003-1724-2922, ORCID 0009-0009-5128-1870, ORCID 0000-0001-8004-3809

Abstract. Virtual reality (VR) has become a key tool in education, although some students still struggle with focus, leading to high dropout rates. The goal of this research was to create a more engaging and effective learning environment through immersive technology, but high costs remain a barrier for low-budget educational institutions. This study evaluates the implementation of an open-source immersive VR (IVR) solution for education, combining OpenSimulator, Firestorm VR, SteamVR and Meta Quest 2, as a cost-effective alternative to commercial platforms. Through systematic hardware benchmarking, we demonstrate a cost reduction compared to commercial solutions while maintaining high performance through optimized hardware configurations. Comparative analysis reveals that open-source solutions offer superior customization and privacy control, although they require greater technical experience for implementation. These findings provide actionable guidelines for institutions adopting IVR, balancing performance and affordability.

Keywords: Immersive virtual reality, Metaverse, Meta Quest 2, OpenSimulator, Open-source, 3D virtual world

1 Introduction

Current education systems confront significant challenges to keep students engaged in digital learning spaces, leading many institutions to explore immersive technologies

like VR. Although VR shows great promise for creating interactive educational experiences, high costs make it difficult for most institutions to implement. Studies show that bringing VR into classrooms requires expensive equipment, specialized software, and teacher training (Radianti et al., 2020). This creates an unfair situation where only well-funded institutions can afford these advanced tools, while others fall further behind, a pattern that research confirms widens existing educational gaps (Pellas et al., 2021).

A significant factor behind the growing interest in immersive technologies is the evolving profile of today's learners. Empirical evidence shows that contemporary students tend to exhibit shorter attention spans, stronger preference for active learning, and higher expectations for interactive digital environments (Makransky and Lilleholt, 2018 ; Parong and Mayer, 2018). Traditional 2D interfaces often fail to provide the level of engagement required to maintain attention and promote deep learning. In contrast, immersive VR (IVR) enhances presence, embodiment, and cognitive focus, offering a promising alternative for addressing attention-related disengagement.

Open-source 3D virtual world such as OpenSimulator represent an opportunity for higher education institutions seeking to adopt immersive technologies without the prohibitive costs typically associated with commercial IVR platforms. These environments allow the creation of large-scale, customizable virtual campuses that can be accessed with low-cost consumer-grade VR headsets.

Despite the potential of IVR technologies, previous research has rarely addressed how institutions with restricted budgets can transition from conventional 2D/3D platforms to fully immersive environments. We examine not only the pedagogical benefits of IVR but also the economic and technical feasibility of deploying open-source VR solutions in resource-constrained institutions.

This work proposes a methodological framework for transitioning from non-immersive and semi-immersive systems to fully immersive virtual reality environments using open-source tools (OpenSimulator, Firestorm VR) and affordable hardware (Meta Quest 2). The framework includes benchmarking of performance parameters, evaluation of cost - effectiveness, and identification of minimum hardware specifications required to ensure fluid and comfortable immersive experiences.

In summary, this study aims to provide a replicable roadmap for higher education institutions to adopt IVR environments using open-source technologies, while addressing three key dimensions: (1) pedagogical alignment with current learner needs, (2) technical performance and hardware feasibility, and (3) cost-effectiveness supported by pricing sources. This approach positions the research as a practical and equitable solution for institutions seeking to modernize their digital learning ecosystems under financial constraints.

1.1 Non-immersive virtual reality to immersive virtual reality

Due to technological limitations and high costs, VR was largely confined to specialized fields such as medical simulation, flight simulators, and industrial design, making it inaccessible to most people. Attempts to popularize VR in the 1990s, such as Nintendo's

Virtual Boy and Sega's VR, ended in commercial failure. However, recent advances in mobile and wearable technologies, now more accessible and affordable, have brought us to what some consider a defining moment for VR (Huggett, 2020).

VR is a technology designed to transport users into digitally constructed environments, commonly referred to as the metaverse. This technology creates a compelling sense of physical presence, defined as the psychological perception of 'being there' within a computer simulated space (Lyu et al., 2023). VR achieves this through experiential interaction with simulated environments, that replicate physical reality using specialized interfaces, including head-mounted displays (HMDs), motion controllers, and haptic feedback systems (Fusaro, 2025). VR systems can be categorized into three distinct modalities based on immersion levels: non-immersive, semi-immersive, and immersive (Fig. 1) (Salatino et al., 2023). Non-immersive, the most basic form, utilizes standard desktop computers where users observe virtual environments through a screen interface and interact via traditional input devices like keyboards or mouse, an approach that generally offers limited user engagement. Semi-immersive systems represent an intermediate level, employing advanced graphics workstations to deliver enhanced spatial presence while maintaining some connection to the physical environment. The highest level, fully immersive, provides complete environmental engagement through HMDs or similar head-coupled devices, creating the most authentic simulation of virtual spaces by surrounding the user's visual field and tracking head movements in real-time. (Ibrahim and Juhari, 2019; Salatino et al., 2023; Stracke et al., 2025). Fully immersive systems also enable spatial and embodied learning, which is particularly beneficial for tasks requiring complex visualization or kinesthetic engagement (Wu et al., 2020).

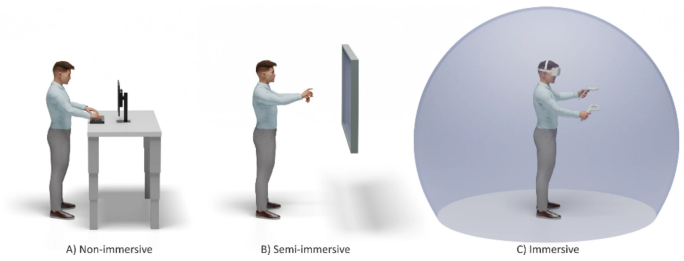


Fig. 1. Non-immersive, semi-immersive and immersive modality (Stracke et al., 2025).

However, the transition toward higher levels of immersion is constrained by financial and infrastructural limitations. High-end HMDs (e.g., Valve Index, HTC Vive Pro) require powerful computers and represent a substantial investment for many institutions.

Unlike conventional virtual platforms or 2D multiplayer online games, IVR systems employ advanced technological components to create a deeply engaging experience. These systems integrate high-precision motion tracking sensors and robust graphics processing units to generate uninterrupted stereoscopic visuals displayed through HMDs (Won et al., 2023). This technical configuration enables real-time synchroniza-

tion between user movements and virtual environment responses, producing a heightened sense of spatial presence unattainable through traditional screen-based interfaces.

In contrast, open-source platforms such as OpenSimulator offer a cost-efficient entry point into immersive learning. When combined with affordable consumer devices such as the Meta Quest 2, institutions can achieve an effective transition from non-immersive or semi-immersive environments to fully immersive VR while minimizing costs and maintaining pedagogical quality.

1.2 Immersive vision and technological transition

The VR market offers diverse HMDs catering to different needs and budgets. Standalone headsets like the Meta Quest 3 (Meta, 2025) provide wireless freedom with integrated processors, while PC-tethered systems such as the Valve Index (Valve Corporation, 2019) deliver high-fidelity graphics for advanced users. Enterprise-focused devices (Varjo XR-4 series) offer industrial-grade precision for professional applications (Varjo, 2023). Recent advancements in mixed reality (Apple Vision Pro) blur boundaries between physical and digital environments (Apple Inc., 2024). Cost-effective options like PICO 4 (PICO Interactive, 2022) compete in the consumer segment, though performance varies significantly across price tiers.

The transition from conventional 2D/3D interfaces to fully immersive experiences with open-source software presents distinct technological and perceptual challenges. The goal is achieving what researchers' term 'unconscious presence', a state where users become so thoroughly immersed that they cease to perceive the virtual environment as artificial (Gonçalves et al, 2021). This phenomenon represents the gold standard for VR systems, yet its attainment requires careful attention to multiple interdependent factors.

Three critical components emerge from the literature as fundamental to this transition. First, high-quality visual representation forms the foundation of immersion, requiring exceptional graphic fidelity, realistic textures, and precise spatial rendering to create convincing virtual spaces (Chessa et al., 2016; Mastrolembo Ventura et al., 2022). Second, accurate positional and motion tracking proves essential, as even minor latency or inaccuracies in head and body tracking can disrupt the illusion of presence and potentially induce simulator sickness (LaViola, 2000). Third, spatially rendered 3D audio significantly enhances immersion when properly synchronized with visual events, with studies demonstrating its role in reinforcing spatial awareness and emotional engagement (Cabero Almenara and Puentes Puente, 2023; Sosa Jiménez et al., 2020).

To systematically examine these requirements, we conducted a comprehensive review of IVR literature, incorporating foundational studies and recent technological advancements (Chessa et al., 2016; Dong and Lee, 2022). This synthesis reveals that achieving unconscious presence depends not merely on individual system components performing optimally, but on their seamless integration and synchronization within the virtual environment.

Overall, immersive vision represents a critical stage in the evolution of digital learning environments. By examining the technical foundations of immersion and aligning them with learner characteristics and institutional constraints this revised section prepares the rationale for the open-source implementation detailed in the methodology.

1.3 Immersive virtual reality for education

As an implementation of VR technology, OpenSimulator emerges as a free 3D virtual world platform specifically designed for educational institutions across all academic levels, from K12 schools to universities. This tool facilitates a paradigm shift in knowledge transfer by integrating VR and immersive technologies to create engaging learning experiences. As Martínez-Gutiérrez et al. (2024) highlight, this tool provides a virtual world where students can freely explore while simultaneously interacting with their teachers, enabling collaborative discovery of new learning horizons. In this context, virtual worlds serve as a complementary, transferable, and effective alternative for enhancing student-instructor interaction and immersive learning experiences.

Open-source platforms such as OpenSimulator provide a viable foundation for educational institutions seeking to deploy IVR without incurring the high costs associated with commercial metaverse solutions. These platforms support customizable virtual campuses, collaborative learning spaces, and multi-user simulations, all while avoiding licensing fees and offering broad interoperability with diverse hardware and software ecosystems.

OpenSimulator distinguishes itself through its modular architecture and high degree of configurability. The platform enables both end-users and developers to configure the software to meet specialized needs, facilitating diverse implementation use cases (Tlili et al., 2022). This versatility makes it especially suitable for applications ranging from pedagogical simulations and vocational training modules to virtual collaboration hubs and interactive entertainment systems (Saenz, 2025).

OpenSimulator's versatile architecture embodies many key characteristics that leading experts associate with metaverse development. Industry leaders have conceptualized the metaverse through complementary perspectives. Zuckerberg (2021) characterizes it as a persistent, 3D digital ecosystem that blends virtual and augmented reality (AR) to enable novel forms of social interaction and shared presence across geographical boundaries. Similarly, Sweeney (2021) of Epic Games describes it as an evolving network of interoperable virtual spaces supported by common technical standards, where users collectively generate and experience digital content across gaming, education, and social domains.

Previous studies demonstrated that IVR environments yield significantly better learning outcomes than traditional instructional methods like lectures or low-immersion media. These comprehensive studies indicate a moderate effect size advantage for IVR across various educational contexts, suggesting its potential as a transformative pedagogical tool (Hamilton et al., 2021; Wu et al., 2020). The increased sensory engagement and spatial presence afforded by IVR systems appear to facilitate knowledge reten-

tion and conceptual understanding more effectively than conventional two-dimensional learning formats.

Despite these advantages, institutions must consider several challenges before integrating IVR into educational programs. Performance bottlenecks, hardware limitations, and connectivity issues can affect the quality of immersive experiences.

Finally, the transition toward IVR is influenced by cost considerations related to hardware acquisition. Devices such as Meta Quest 2 provide an affordable entry point, with publicly documented pricing available from Meta's official platform (Meta Support, 2025). In contrast, enterprise-oriented HMDs like the Varjo XR-4 exceed USD 3,000 and require high-end workstation hardware.

1.4 Study objectives

This study aims to develop and document a reproducible implementation framework for adapting open-source 3D virtual worlds (OpenSimulator) to IVR in the educational field, using consumer hardware (Meta Quest 2) and modified open-source viewers (Firestorm VR).The proposed framework addresses both pedagogical and technological needs, ensuring alignment with the characteristics of contemporary learners as well as the financial limitations of higher education institutions.

The research will pursue three key technical objectives to achieve this goal:

- **O1**: To evaluate the cost-saving potential of this open-source approach by comparing total expenses (hardware and software) against commercial VR education solutions
- **O2**: To identify and optimize technical bottlenecks in the implementation process (e.g., latency, visual rendering settings, wireless streaming) that affect educational usability.
- **O3**: To establish best practices for deploying this solution in resource-constrained institutions, including minimal hardware requirements and scalability considerations.

2 Methodology

This study employed a mixed-methods research design combining quantitative hardware benchmarking with qualitative usability analysis to evaluate the feasibility of implementing an open-source immersive virtual reality (IVR) ecosystem, focusing on three core components: (1) OpenSimulator as the virtual world platform, (2) Firestorm VR as the optimized viewer, and (3) Meta Quest 2 as the target HMD. Performance metrics - including frames per second (FPS) and latency - were systematically recorded across all hardware configurations.

As an open-source alternative to Second Life, OpenSimulator provides a robust framework for developing and managing 3D virtual environments. The platform of-

fers comprehensive capabilities for building virtual spaces, customizing environmental elements, and facilitating real-time interaction among multiple users within these simulated settings. Its architecture supports the creation of immersive educational environments while maintaining flexibility for technical adaptation and pedagogical integration. OpenSimulator provides 3D visualization through conventional LCD monitors, with user interaction primarily mediated through keyboard-based controls. This study investigates methods to enhance interaction quality in virtual environments through advanced immersion techniques.

The central research question examines: How can we effectively transition users from conventional 2D/3D displays to fully IVR experiences while maximizing the sense of physical presence and natural interaction? This research question focuses on overcoming the technical and perceptual challenges of creating seamless immersion, where users wearing VR headsets (Meta Quest 2 in this study) can achieve authentic spatial awareness and intuitive control. Currently, newer versions of the headset are available, however, for this research, the Meta Quest 2 was used due to current availability, although the methodology is adaptable for future HMD updates. Taking into account hardware benchmarking and the feasibility of implementing a low-cost immersive virtual reality (IVR) ecosystem, a comparative table of IVR-compatible software is shown in Table 1.

Table 1. Comparative analysis of VR/3D world software

Software	**License**	**IVR Compatibility**	**Suitability for Education**
OpenSimulator	Open-source	Via Firestorm VR Mod	Highly flexible and scalable; ideal for low-budget institutions (OpenSimulator, 2024).
Second Life	Free / Commercial	Limited (no native IVR)	Useful for exploration but not optimal for IVR teaching (Linden Lab., 2024).
Mozilla Hubs	Open-source	Partial (WebXR)	Highly accessible; limited graphic fidelity (Mozilla, 2024).
ENGAGE XR	Commercial	Native IVR	Strong for enterprise training; expensive for large classes (Engage XR, 2025).
ClassVR / Eduverse	Commercial	Native IVR	Good plug-and-play solution but costly and proprietary (ClassVR, 2025).
Unity + VR Toolkit	Free tier / Paid	Native IVR	Excellent for custom experiences; requires advanced skills (Unity Technologies, 2024).

A secondary question investigates: What minimal hardware specifications (GPU/ CPU/ RAM) are required to maintain acceptable performance in OpenSimulator-based IVR learning environments? For an optimal and comfortable IVR experience, 90 FPS is widely considered the minimum acceptable standard, with 120 FPS or higher being ideal to minimize latency, motion sickness, and disorientation (Cobb et al., 2019).

Below 90 FPS, users may experience cybersickness, lag, and reduced presence (Rebenitsch & Owen, 2016). Although FPS is largely limited by video processing capabilities, a comparison of VR headsets to analyze the display FOV and some of their pros and cons can be viewed in Table 2.

Table 2. Comparative analysis of representative VR headsets

Device	Display / FOV	Cost	Pros	Cons
Meta Quest 2	1832×1920 per eye ∼100°	USD $249 (Meta, 2025)	Affordable, wireless, easy to deploy.	Requires PC link for high-end IVR.
Valve Index	1440×1600 per eye 130°	USD $999 (Valve Corporation, 2019)	Excellent fidelity, wide FOV.	High cost; requires powerful PC.
HTC Vive Pro 2	2448×2448 per eye	USD $799 (HTC Corporation, 2024)	Extremely high resolution.	Expensive; complex setup.
PICO 4	2160×2160 per eye	USD $429 (PICO Interactive, 2022)	Good balance price/quality.	Limited availability in some regions.
Varjo XR-4	2880×2720 per eye	USD $3,990 (Varjo, 2023)	Best-in-class fidelity.	Prohibitively expensive for most institutions.

This inquiry, along with parallel investigations into cost-effective adaptation methods, directly supported the development of an implementation framework for converting open-source 3D virtual worlds to IVR educational environments.

2.1 OpenSimulator configuration

As an open-source alternative to Second Life, OpenSimulator was selected due to its cost-effectiveness, modular architecture, and extensive configurability for multi-user educational environments. Although is natively designed for Windows systems, this implementation was deployed on a Linux Debian 12 platform (kernel 5.10.0-33) to leverage its superior memory management capabilities. The adaptation process necessitated the Mono Project's compatibility layer to bridge the Windows-native OpenSimulator with the Linux environment, ensuring stable operation while optimizing resource allocation.

Mono is an open-source implementation of the .NET framework, designed to allow developers to create applications across platforms. In this study, Mono was installed via Debian's official repositories, as shown in Fig. 2.

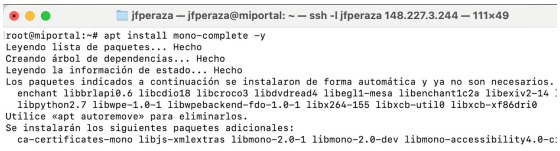


Fig. 2. Mono's project installation.

Version 6.8 of the Mono project was used, with SUID permissions enabled for execution via bash shell from the command line, as shown in Fig. 3.

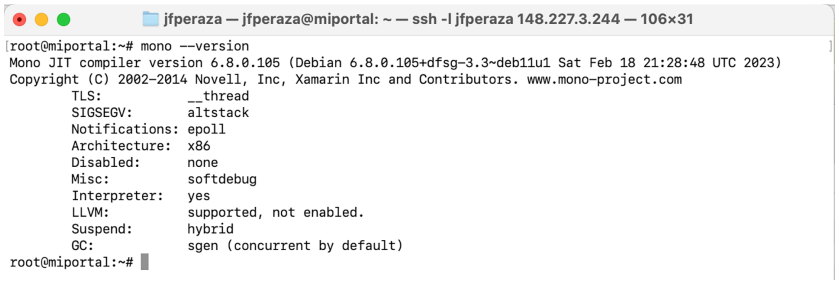


Fig. 3. Mono command line version and execution.

Once the installation process of the Mono Project was completed, a documentary review was conducted to determine which database management system would best meet the requirements of this implementation. Both PostgreSQL and MySQL were evaluated, considering factors such as stability, resource optimization, and availability of documentation.

Although PostgreSQL exhibited superior performance in benchmarked parameters, MariaDB (an open-source MySQL derivative) was selected due to its more efficient memory management, comprehensive documentation, and robust community support (Zapata, 2024). While marginally less performant, these practical advantages aligned better with project requirements. Fig. 4 illustrates the database creation process in MariaDB.

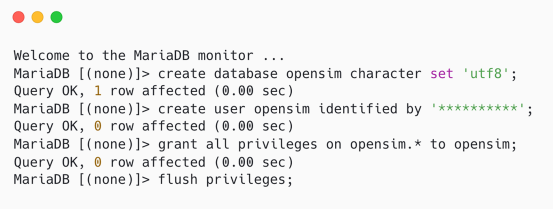


Fig. 4. Database creation in MariaDB.

The OpenSimulator platform was installed according to the following procedure: First, a directory was created within the home folder of a non-administrative user account. The latest stable release (v4.2) was obtained from the official GitHub repository. After extracting the compressed files, the source code directory was accessed, and the compilation process was executed using default build parameters (Olivarría González et al., 2023). This installation approach was adopted to preserve system security while guaranteeing compatibility with the experimental environment. The compilation process is illustrated in Fig 5.



Fig. 5. Compiling OpenSimulator.

After the installation process, the initial virtual world configuration was implemented through modification of the 'MyWorld.ini' configuration file. This critical setup phase established the foundational parameters for the virtual environment. The configuration procedure and resulting parameters are documented in Fig. 6, which presents the complete set of implemented settings and their organizational structure.

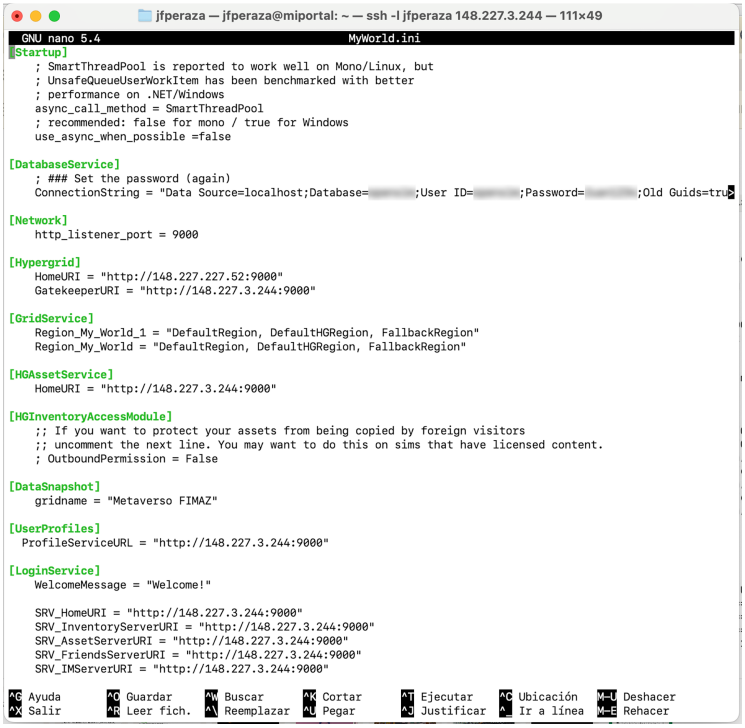


Fig. 6. First virtual world setup.

The virtual world deployment began with initialization of the core OpenSimulator system. A background session was established using the GNU screen terminal multiplexer, followed by execution of the metaverse environment through specific command-line instructions. The complete command sequence and system response were documented in Fig. 7, showing the startup protocol.

```
opensim@gnu:~$ screen -S opensim
opensim@gnu:~$ mono OpenSim.exe
```

Fig. 7. Command sequence to start OpenSimulator.

Finally, OpenSimulator environment was accessed through a web browser by entering the server's IP address and the port configured in the MyWorld.ini file. For web-based administration, the system utilized port 9000, forming the access URL: http://serverIP:9000/wifi/. Fig. 8 illustrates the browser interface used for metaverse configuration, displaying the administrative dashboard and connectivity settings.

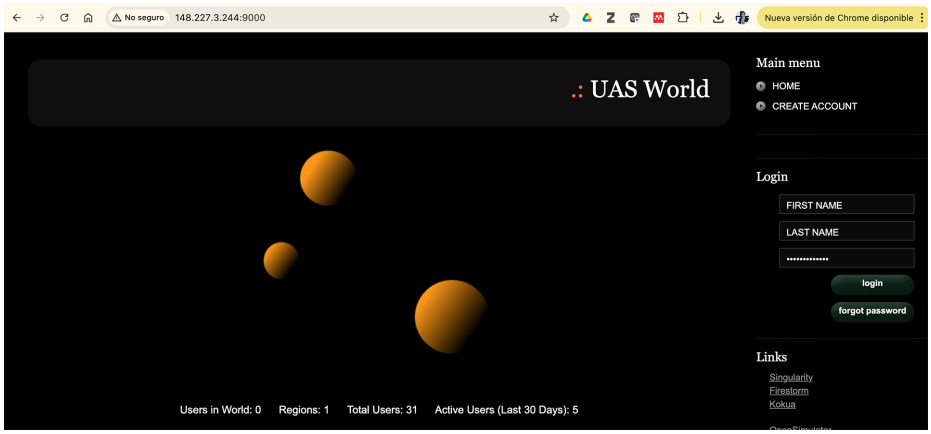


Fig. 8. Metaverse web administration interface.

Following the metaverse configuration and user account creation, the connection was established from a client PC using Firestorm Viewer (v6.6.9), a widely adopted open-source client for virtual world access. The compatible version was downloaded according to the client's operating system specifications (Windows 11 or Ubuntu 22.04 LTS). As noted by Lyon (2024), Firestorm serves as the gateway to virtual environments, analogous to web browsers for Internet access. Fig. 9 documents the authentication interface displayed during login to the virtual world created 'UAS World' using the test account 'Juan Peraza'.

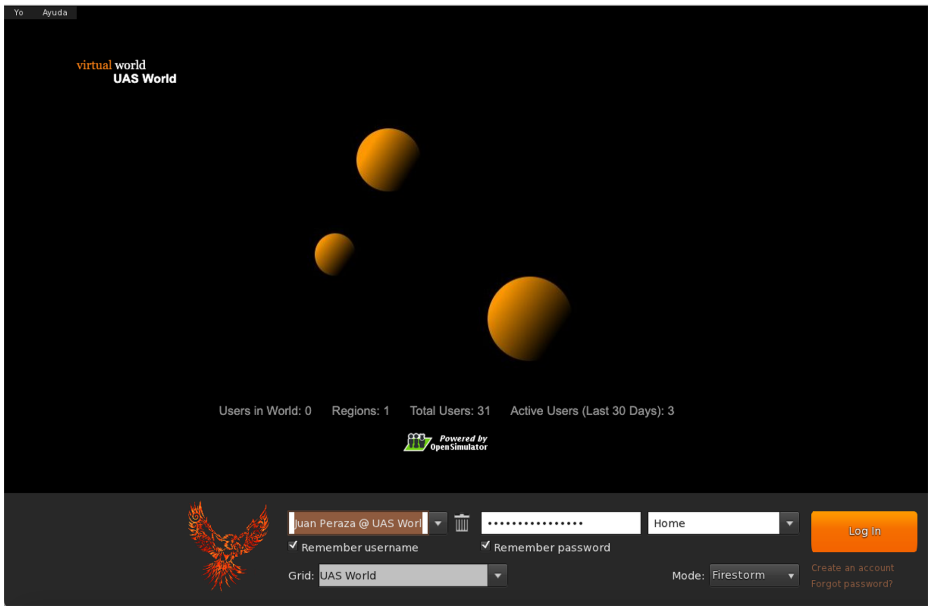


Fig. 9. Firestorm login interface.

After successful login through the Firestorm viewer, the access to the UAS World virtual environment was established. Fig. 10 presents the customized avatar within the metaverse, displaying the preconfigured clothing options and modified physical characteristics that were implemented for user representation.



Fig. 10. Customized avatar appearance in UAS World virtual environment.

2.2 Immersive VR Integration

After establishing OpenSimulator's operational status, the system configuration proceeded with the integration of immersive visualization components. The Meta Quest 2 headset was connected through Virtual Desktop software to enable VR functionality. This hardware-software combination facilitated the transformation of the virtual environment into a fully immersive experience.

2.2.1 Meta Quest 2. The study selected pre-existing Meta Quest 2 headsets as the immersive interface. The devices' ergonomic design, high-resolution display, and submillimeter motion tracking precision (Fig. 11) were identified as critical features for sustained educational use. Unlike previous models, this standalone headset required no external PC connection, offering enhanced portability and accessibility (Meta, 2025). These technical characteristics made it particularly suitable for immersive virtual environment applications.

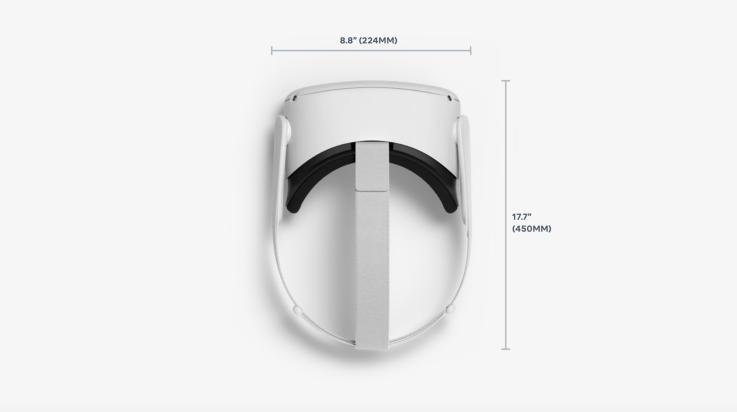


Fig. 11. Meta Quest version 2 (Meta, 2025).

The Meta Quest 2 employed an integrated camera-based tracking system to monitor head and hand movements, facilitating natural interaction within virtual environments. As a standalone device requiring no external computer tethering, it offered enhanced portability while maintaining stable performance. These technical capabilities, combined with access to an extensive application ecosystem (Meta, 2025), positioned the headset as an optimal solution for diverse VR applications, including gaming, social interaction, and immersive educational experiences.

2.2.2 Virtual desktop.

The virtual desktop implementation utilized a cloud-based desktop virtualization service, which was achieved through software emulation on remote virtual devices. This solution was specifically designed for teleworking applications, enabling remote network access to documents and computing resources. Prior to deploying the virtual

desktop infrastructure, an assessment of user requirements was conducted, and the network infrastructure was prepared. The implemented system supported wireless connectivity to multiple computers, facilitating various applications including media viewing, Internet browsing, gaming on virtual displays, and PC VR game streaming. The virtual desktop application was selected for its optimized performance characteristics, featuring low-latency operation and high-quality streaming capabilities (Canorea, 2025).

2.2.3 Connecting Meta Quest 2 to the computer.

The connection between devices was established using either a USB-C data cable or wireless transmission. Both methods were implemented, though wireless connectivity was prioritized to eliminate potential safety hazards and mobility restrictions associated with physical cabling. Following installation of required applications, network configuration was performed via a local area network (LAN), ensuring all devices operated on the same network. The Meta Quest 2 headset was specifically connected to a Wi-Fi 6 network, achieving transmission speeds up to 9.6 Gbps. This high-speed connection enhanced FPS rates and improved visual rendering quality during operation.

2.2.4 SteamVR and Firestorm VR setup.

The VR implementation began with SteamVR, described by its developers as a comprehensive platform for VR content across multiple hardware systems, including Valve Index, HTC Vive, Oculus Rift, and Windows Mixed Reality headsets (Valve Corporation, 2025). Following installation on a Windows 11 system, the specialized Firestorm VR (version 6.6.8 + VR Mod 6.3.3) was configured. This modified viewer, adapted for Second Life and OpenSimulator compatibility with HMDs, employed SteamVR for system integration.

The VR implementation followed a minimalistic design philosophy to reduce maintenance requirements and enhance long-term viability. Successful operation required installation of both Meta Quest 2 and SteamVR drivers. Upon launch, Firestorm VR automatically initialized all necessary SteamVR subsystems for headset functionality.

The VR capabilities described are based on a dual-screen split of the primary display, which is essential for VR headset compatibility. One screen is rendered per eye, helping to create the immersive experience, as shown in Fig. 12. This feature is not supported by the standard Firestorm client, hence the necessity of the VR version.

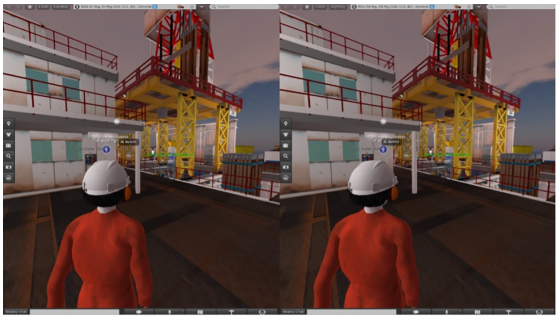


Fig. 12. View from inside Firestorm VR (Austin, 2022).

2.2.5 Firestorm VR and Meta Quest 2 configuration parameters.

Following device connection, multiple visual parameters required configuration within the Firestorm VR interface to optimize display performance through the Meta Quest 2 headset. Initial setup configurations frequently exhibited technical challenges including improper distance scaling, focus inaccuracies, and incorrect lens separation parameters. These issues were systematically addressed through iterative calibration of the VR rendering settings. Key parameters to configure include:

- Interpupillary Distance (IPD): This refers to the distance, in millimeters, between the user's pupils. For optimal image clarity, the lens spacing should closely match the user's IPD. Meta Quest 2 headsets are best suited for IPD values ranging from 56 mm to 70 mm, which corresponds to approximately 95% of the adult population (Meta, 2025).
- Focus distance: This setting adjusts the focal distance for image clarity.
- Texture shift and texture zoom: These parameters enhance image texture quality.
- Field of View (FOV): In extended reality (XR), FOV refers to the extent of the virtual or augmented environment visible through the headset at a given moment.

The FOV was identified as a critical parameter influencing UX in immersive environments. In the VR implementation, the expanded FOV enhanced spatial awareness and immersion through increased peripheral content visibility (Fig. 13). By contrast, the AR system operated with a narrower FOV, which constrained the superimposition area for digital elements onto physical environments. These specifications aligned with established guidelines from the Interaction Design Foundation (2023), with the VR configuration specifically optimized to maintain the recommended 90°-110°range for optimal performance.



Fig. 13. Ideal FOV area in VR (Interaction Design Foundation, 2023).

When designing immersive experiences, you'll find VR and AR require fundamentally different approaches to FOV optimization. In VR, you should target a wide 90°-110° FOV to maximize presence but must carefully balance this with comfort considerations, avoiding extreme peripheral motion that could trigger simulator sickness. The expanded viewport lets you create enveloping environments, though rapid movements near the edges may disorient users. Conversely, AR's narrower 30°-50° FOV

forces you to work within tighter constraints (Interaction Design Foundation, 2023). Here, you need to anchor key content centrally (Fig. 14) and maintain clear visibility of all interactive elements, as users can't rely on peripheral awareness like in VR. While VR allows you to construct entire worlds, AR demands you strategically layer digital elements onto physical spaces without requiring excessive head movement.

Both mediums share the need for thoughtful content placement, but where VR pushes you to expand the view, AR challenges you to make every degree count through precise spatial composition and context-aware design.

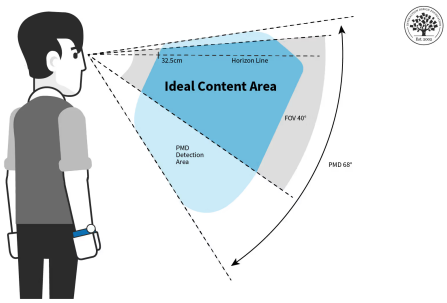


Fig. 14. FOV restricted in AR (Interaction Design Foundation, 2023).

Based on the established technical parameters, the Meta Quest 2 headset was configured with the following optimized settings, as documented in Fig. 15: IPD of 68 mm, focus distance set to 10 units, texture shift adjusted to 18.5%, texture zoom at 79%, and FOV fixed at 100°. These calibrated values represented the optimal balance between visual comfort and immersive quality for the experimental conditions.

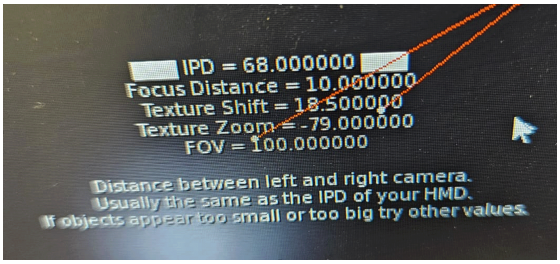


Fig. 15. Firestorm VR configuration.

The final configuration was specifically calibrated for an adult test subject (age 42 years), verifying that the configured IPD falls within the normal range for adults (Meta, 2025). This age-specific calibration ensured optimal visual alignment and comfort during extended use sessions.

Continuing with the immersive vision conversion process, and after successfully setting up OpenSimulator, Firestorm VR, and configuring the Meta Quest 2 headset, the next step involves establishing the connection between SteamVR and Firestorm

VR. It is worth recalling that Firestorm VR is a customized viewer derived from the Firestorm client, while SteamVR is a VR module developed by Steam, both installed on a PC running Windows 11.

To establish connectivity between Firestorm VR and SteamVR, the following procedure must be executed: First, launch Firestorm VR on the Windows 11 system. Then, press CTRL+TAB to activate or deactivate the SteamVR driver. This operation is required every time the user initiates VR mode following system startup.

2.3 Basic avatar usage within the metaverse with immersive vision enabled

Traditional avatar navigation in the virtual environment presented significant complexity, requiring keyboard inputs for movement and mouse control for view orientation. The implementation of immersive vision technology substantially improved this interaction paradigm by introducing natural head-tracking capabilities, thereby eliminating dependence on manual view direction adjustment. Furthermore, the system was enhanced through integration of an Xbox game controller connected via Bluetooth, as documented in Fig. 16. This peripheral device provided more intuitive avatar movement control compared to standard keyboard input methods.



Fig. 16. Xbox control (Microsoft, 2025).

3 Results

The transition from conventional open-source 3D virtual world into an immersive experience was successfully implemented through the installation and configuration of specialized tools. The core components included: OpenSimulator as the virtual world platform, Firestorm VR as the optimized viewer for VR environments, and SteamVR as the compatibility module for immersive device integration, specifically adapted for the Meta Quest 2 headset. This integrated system enabled full transformation of conventional 3D content into immersive educational experiences while maintaining open-source accessibility.

3.1 Hardware configuration

The results of the hardware benchmarking, immersive configuration, and cost analysis support the feasibility of this approach for resource-limited higher education institutions. Regarding the headsets, Table 3 compares standalone, PC-tethered, and enterprise VR headsets based on representative models, principal uses, and price ranges with data sourced from manufacturer specifications. It is important to note that the prices shown in this section were obtained directly from the provider's website, Amazon México or Amazon US.

Table 3. Key features of major VR headset categories (Meta, 2025; Valve Corporation, 2019; Varjo, 2023)

Type	Example	Best For	Price Range	Source
Standalone	Meta Quest 2/3	Consumers / Education	USD $300–$500	Amazon MX
PC-Tethered	Valve Index	Enthusiasts / Developers	USD $600–$1,000	Provider's Web
Enterprise	Varjo XR-4	Professional Training	USD $3,000+	Provider's Web

The study prioritized server-side hardware specifications for hosting the core OpenSimulator metaverse processes, distinguishing them from client-side requirements for Firestorm VR operation. Notably, client machines demanded superior hardware capacity due to local image rendering responsibilities, necessitating robust performance for optimal visual fluidity.

Initial tests with Intel i3 and i5 server processors yielded suboptimal frame rates ($\leq$ 20 FPS). Subsequent implementation of an Intel i7 processor demonstrated significant improvement, achieving peak performance of 237 FPS. Table 4 details the optimized production server specifications deployed in the final implementation.

Table 4. Minimum optimal hardware requirements for OpenSimulator server operation

Hardware	Description
CPU Processor	Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz
Graphics Card	Integrated
GPU	Coffee Lake GT2, 192 cores
RAM Memory	8 GB
Hard Drive	512 GB SSD

Client-side hardware requirements were more demanding due to the client-dependent rendering architecture, which shifted most graphical processing from the server to local machines. The project evaluated three hardware configurations: An initial test system with an Intel i5 processor (8GB RAM) and GTX 1660 Super GPU (6GB VRAM, 1408 CUDA cores) exhibited critical performance limitations, failing to maintain stable connectivity between SteamVR and Firestorm VR.

Subsequent testing with an equivalent Intel i7 configuration replicated these interoperability issues. Ultimately, a high-performance workstation with a 14th-generation Intel i9 processor (16 cores, no integrated graphics) and dedicated Gigabyte PCI-Express GPU resolved the bottlenecks. Complete specifications are detailed in Table 5.

Table 5. Client-side hardware configurations evaluated

Component	Config A	Config B	Config C
CPU Processor	Intel Core i5-6500 4 cores, 3.6 GHz	Intel Core i7-7700 4 cores, 5.1 GHz	Intel Core i9-12900F 16 cores, 5.1 GHz
Graphics Card	NVIDIA GTX 1660 Super 6 GB VRAM, PCIe 4.0	NVIDIA GTX 1660 Super 6 GB VRAM, PCIe 4.0	Gigabyte Radeon RX 6600 8 GB GDDR6, PCIe 4.0
GPU Cores	1408 stream processors	1408 stream processors	1792 stream processors
System Memory	8 GB DDR3 RAM	8 GB DDR3 RAM	32 GB DDR4 RAM
Storage	512 GB SSD	512 GB SSD	1 TB NVMe SSD

3.2 Inmersive Rendering and Quality

The 237 FPS shown in Fig. 17 (client-side Firestorm testing) were achieved using the "Config C" hardware configuration for client, as previously specified.

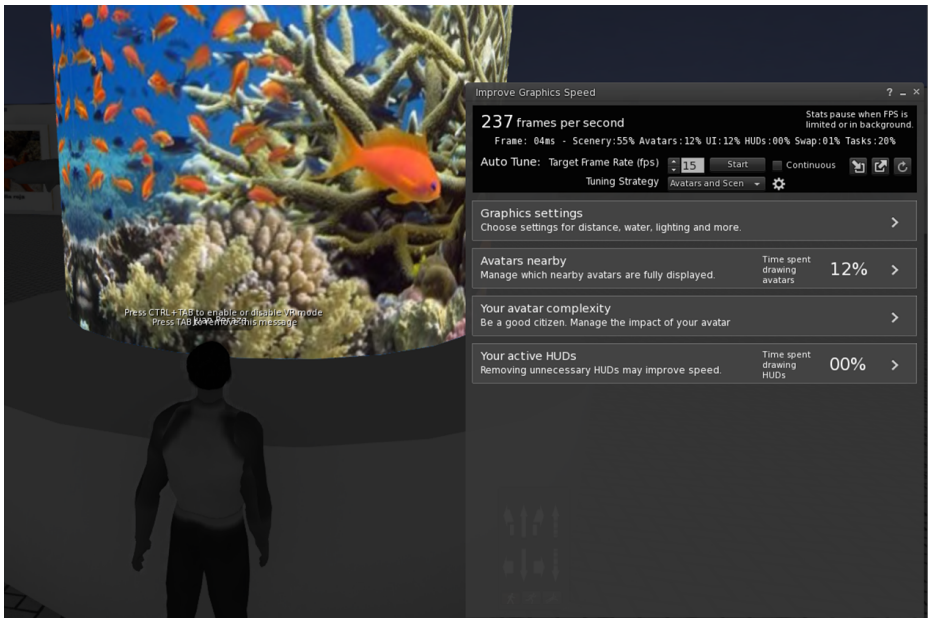


Fig. 17. FPS results before implementing immersive vision.

The implementation process involved three key technical adaptations: graphics parameter optimization, IPD calibration, and FOV adjustment. These modifications enabled seamless integration into the platform and fluid visual performance. System latency was maintained at optimal levels thanks to Wi-Fi 6 network connectivity between the headset and the host computer, enabling seamless interaction with the environment in real time.

Fig. 18 shows the performance results comparing two configurations, SteamVR with Firestorm VR achieved 90 FPS on the Meta Quest 2, versus 237 FPS in Firestorm's non-immersive mode.

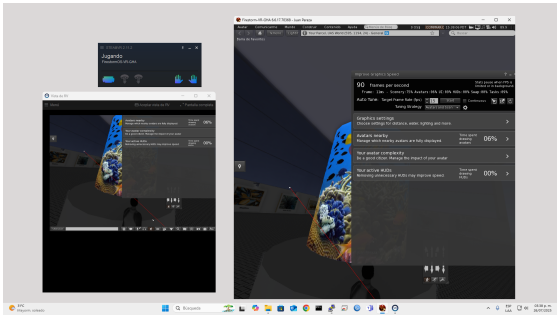


Fig. 18. FPS results before implementing immersive vision.

The resulting configuration demonstrated full functionality for immersive OpenSimulator visualization, with head-tracking responsiveness that met all performance benchmarks. Stability testing confirmed consistent performance during extended sessions ($\geq$ 2 hours continuous use), with no significant frame rate drops or system crashes. These results validate the technical feasibility of transforming conventional open-source 3D environments into immersive experiences, establishing a foundation for educational applications requiring interactive learning frameworks.

3.3 Cost Analysis and Price Sources

As institutions seek to modernize pedagogical approaches and prepare for the digital future of education, evaluating the financial feasibility and scalability of VR implementations becomes imperative. Table 6 describes a cost analysis of commercial and open-source IVR platforms tailored for educational purposes. The commercial IVR platforms examined offer turnkey or subscription-based solutions that include pre-packaged content, technical support, and varying degrees of customization. In contrast, open-source alternatives such as OpenSimulator paired with Meta Quest 2 present a highly adaptable and cost-effective model but in addition to HMDs, it requires a PC and greater technical skills for configuration and maintenance.

It should be noted that some platforms, including Class VR and ENGAGE XR, impose minimum purchase requirements for either licenses or hardware equipment, which may affect their accessibility. Unlike custom solutions, commercial platforms prioritize plug-and-play usability and low technical knowledge requirements. However, their cumulative costs can become significant over time, particularly when scaling to larger classrooms or entire campuses. In contrast, open-source solutions, notably those based on OpenSimulator paired with Meta Quest 2 (or similar headsets), offer a low-cost, high-flexibility alternative. The software itself is free to download and deploy, with the primary investment being hardware and the technical expertise required for configuration and maintenance of the virtual environment.

Table 6. Commercial vs. Open-Source IVR solutions for education

Platform	Estimated Minimum Hardware Cost	Educational License/ Subscription	Estimated Minimum Annual Cost
ClassVR (Stemfinity, 2025)	˜USD $4,900 (8 units)	Eduverse Portal license: USD $769/year	˜USD $5,700 (hardware + license)
Wonda VR (Wonda VR, 2024)	˜USD $300 per headset	Free trial; Basic plan: USD $350/month	˜USD $4,500 (hardware + license)
ENGAGE XR (ENGAGE XR, 2025)	˜USD $300 per headset	EDU License: USD $190/user/year	˜USD $1,250 (hardware + license)
Bodyswaps (Bodyswaps, 2025)	˜USD $300 per headset	Free trial; EDU license: USD $80/user/year (100–250 users)	˜USD $380 (hardware + license)
Open-Source (OpenSimulator, 2024)	˜USD $700 per station	Free (open-source software)	˜USD $700 (hardware)

Table 7 presents a comparative analysis between the average cost of commercial Immersive Virtual Reality (IVR) solutions and a free, open-source ecosystem based on OpenSimulator. The results indicate that the adoption of open-source IVR reduces the initial investment by approximately 60–70% in contrast to commercial alternatives. This cost differential demonstrates that open-source IVR environments constitute a financially viable pathway for institutions with limited budgets, significantly enhancing the accessibility and scalability of immersive learning implementations.

Table 7. Cost Analysis and Price Sources

Solutions	Hardware Cost	Software License Cost	Sources
Commercial IVR Solutions	USD $4,900 to $7,000	˜USD $300 to $1,200	ClassVR (2025) ENGAGE XR (2025) Bodyswaps (2025)
Open-Source IVR Solution	USD $1,400 to $1,700	Free	OpenSimulator Firestorm VR SteamVR

4 Discussion

The results of this study provide compelling evidence that open-source immersive virtual reality (IVR) environments can serve as effective and affordable platforms for enhancing student engagement and learning outcomes in higher education. The key findings performance benchmarking, cost analysis, and comparative software/hardware evaluation must be interpreted in relation to both, the pedagogical issues introduced earlier and the technological constraints facing institutions.

Students tend to exhibit reduced attention spans, require high interactivity, and respond positively to multisensory stimuli. The stable head-tracking, high frame rates, and improved embodiment achieved in the final hardware configuration (Intel i9 + Radeon RX 6600 + Meta Quest 2) directly contribute to sustaining attention and reducing cognitive fatigue, critical factors for today's learners and in the technological

field this configuration demonstrated a clear performance hierarchy maintaning a stable SteamVR-Firestorm VR integration. This aligns with prior research by (Radianti et al. 2020) indicating that client-side rendering in virtual environments demands robust single-threaded CPU performance (evidenced by the i9's 5.1GHz boost) and dedicated GPU resources (1792 stream processors). Recurrent bottlenecks in i5/i7 configurations with identical GPUs demonstrate that performance in OpenSimulator-based IVR environments is constrained not only by graphical capacity but also by CPU processing capabilities.

The performance analysis also showed that underpowered systems (Configurations A and B) produced low frame rates, high latency, and visual instability. These findings align with the literature indicating that suboptimal VR performance can result in increased cognitive load, reduced presence, and higher susceptibility to cybersickness (Cobb et al., 2019).

The cost analysis demonstrates that open-source IVR ecosystems present a financially viable alternative to commercial solutions. However, cost savings do not come without trade-offs. Open-source implementations require more technical expertise to configure and maintain compared to commercial ecosystems. Institutions must therefore evaluate the balance between affordability and the internal technical capacity needed for long-term sustainability.

While the optimal i9/RX 6600 setup incurred higher initial costs, its 237 FPS output justified the investment for educational deployments requiring prolonged use. This finding contradicts common assumptions that GPU selection alone determines VR performance, as shown when identical GTX 1660 Super GPUs underperformed across different CPU tiers. The 32GB RAM and NVMe SSD likely contributed to texture streaming efficiency (Pellas et al., 2021).

For institutions adopting open-source VR solutions, these results advocate prioritizing: (1) latest-generation CPUs with high clock speeds ($>$ 5GHz) over core count alone, (2) mid-range GPUs with at least 8GB VRAM, and (3) fast storage to mitigate asset loading delays.

5 Conclusion

This study demonstrates the technical and pedagogical feasibility of implementing a cost-effective immersive virtual reality (IVR) ecosystem for higher education using open-source software and affordable consumer hardware. By integrating OpenSimulator, Firestorm VR, SteamVR, and the Meta Quest 2, the research provides a viable pathway for institutions with limited financial resources to transition from traditional 2D/3D interfaces to fully immersive learning environments. From a pedagogical perspective, the study confirms that IVR environments enhance engagement, embodiment, and spatial understanding, consistent with prior empirical evidence (Hamilton et al., 2021; Makransky and Lilleholt, 2018).

The implementation of this open-source ecosystem demonstrated the technical feasibility of cost-effective immersive education systems. Performance benchmarking re-

vealed that client-side hardware specifications, particularly high-clock-speed CPUs ($\geq$ 5.1GHz) and dedicated mid-range GPUs ($\geq$ 8GB VRAM), were critical determinants of system stability, with the Intel i9/RX 6600 configuration achieving optimal frame rates (90 FPS) for educational applications. These findings provide concrete hardware guidelines for institutions seeking to adopt VR while minimizing infrastructure costs.

From a technical perspective, integrating the platforms, along with precise configuration of parameters such as IPD, FOV, and other graphical settings, allowed the environment to be tailored to the user's physiological and perceptual characteristics. This technical refinement ensured a fluid and comfortable experience, minimizing the risks of eye strain, disorientation, or cybersickness, issues commonly reported in poorly configured VR environments. Furthermore, the choice of a wireless connection between the Meta Quest 2 headset and the computer contributed to increased user mobility and comfort, reducing physical barriers to prolonged interaction.

The use of natural-head-movement-controlled avatars and Bluetooth devices such as the Xbox Controller was realized in order to provide the user with a more natural and realistic impression. This configuration facilitated a more natural involvement of students with educational tasks, without the extra cognitive load of using physical keyboards or touch screens. Accessibility of the technology and customization to fit individuals' needs were thus important to achieve the full educational potential of the experience in this respect.

This research indicates that while commercial IVR setups deliver optimal performance, their high cost makes them impractical for budget-constrained institutions. In contrast, the proposed OpenSimulator-based solution with Meta Quest 2, cuts initial costs by approximately 60-70% compared to professional alternatives, without significantly compromising immersive quality. However, this cost reduction requires investment in technical staff for setup and maintenance, which must be factored into cost-benefit analyses. This cost-effectiveness positions open-source VR environments as a sustainable solution for higher education, particularly where scalability and accessibility are priorities.

Overall, the research provides a replicable and scalable model for higher education institutions seeking to adopt immersive technologies in a financially responsible manner. While technical barriers and infrastructure requirements remain, open-source IVR ecosystems represent a promising pathway for democratizing access to immersive learning.

The comparative hardware analysis revealed that clock speeds of both the CPU and GPU significantly impacted OpenSimulator's performance. This work establishes a replicable framework for deploying open-source VR education platforms, balancing performance and affordability. Future research directions should investigate: (1) cloud-rendering alternatives to further reduce client hardware demands, (2) scalability in multi-user learning scenarios ($>$ 20concurrent avatars), (3) adaptive content optimization techniques for heterogeneous student devices, and (4) longitudinal academic outcomes and multidisciplinary implementations to fully realize this technology's transformative potential in education.

References

Apple Inc. (2024). *Vision Pro technical specifications*. https://www.apple.com

Austin, A. (2022). *Firestorm VR mod 6.6.3* [Blog post]. AI Austin. https://aiaustin.wordpress.com/2022/09/18/firestorm-vr-mod-6-6-3

Bodyswaps. (2025). *Pricing*. Retrieved August 3, 2025, from https://bodyswaps.co/pricing

Cabero Almenara, J., Puentes Puente, A. (2023). La realidad aumentada: Tecnología emergente para la sociedad del aprendizaje. *AULA Revista de Humanidades y Ciencias Sociales*, 66(2), 35–51.

Canorea, E. (2025). *Virtual Desktop Infrastructure*. Plain Concepts. https://www.plainconcepts.com/es/virtual-desktop

Chessa, M., Maiello, G., Borsari, A., Bex, P. J. (2016). The perceptual quality of the Oculus Rift for immersive virtual reality. *Human–Computer Interaction*, 34(1), 51–82. https://doi.org/10.1080/07370024.2016.1243478

ClassVR. (2025). *ClassVR & Eduverse platform: Educational VR ecosystem*. Retrieved from https://www.classvr.com

Cobb, S. V. G., Nichols, S., Ramsey, A., Wilson, J. R. (2019). Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments*, 8(2), 169–186. https://doi.org/10.1162/105474699566152

Dong, H., Lee, J. (2022). The metaverse from a multimedia communications perspective. *IEEE MultiMedia*, 29(4), 123–127. https://doi.org/10.1109/MMUL.2022.3217627

ENGAGE XR. (2025). *Pricing*. Retrieved August 3, 2025, from https://engagexr.com

Fusaro, M., Lisi, M. P., Era, V., Porciello, G., Candidi, M., Aglioti, S. M., Boukarras, S. (2025). The transformative power of virtual reality. *Topoi*. Advance online publication. https://doi.org/10.1007/s11245-025-10216-1

Gonçalves, G., Monteiro, P., Coelho, H., Melo, M., Bessa, M. (2021). Systematic review on realism in immersive VR/AR/MR. *IEEE Access*, 9, 89150–89161. https://doi.org/10.1109/ACCESS.2021.3089946

Hamilton, D., McKechnie, J., Edgerton, E., Wilson, C. (2021). Immersive VR as a pedagogical tool: A systematic review. *Journal of Computers in Education*, 8(1), 1–32. https://doi.org/10.1007/s40692-020-00169-2

HTC Corporation. (2024). *VIVE Pro 2: Technical specifications*. https://www.vive.com/us/product/vive-pro2/overview

Huggett, J. (2020). Virtually real or really virtual: Towards a heritage metaverse. *Studies in Digital Heritage*, 4(1), 1–15. https://doi.org/10.14434/sdh.v4i1.26218

Ibrahim, E. N. M., Juhari, S. (2019). A preliminary study on 360° IVR vs non-IVR event marketing. *IJATCSE*, 8(1.6), 9–13. https://doi.org/10.30534/ijatcse/2019/0281.62019

Interaction Design Foundation. (2023). What is field of view (FOV) in extended reality? https://www.interaction-design.org/literature/topics/field-of-view-fov-in-extended-reality

LaViola, J. J. Jr. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1), 47–56. https://doi.org/10.1145/333329.333344

Linden Lab. (2024). *Second Life: Virtual world documentation*. Retrieved from https://secondlife.com

Lyon, J. (2024). *Firestorm Viewer*. https://www.firestormviewer.org

Lyu, K., Brambilla, A., Globa, A., de Dear, R. (2023). An immersive multisensory VR approach to studying built environments. *Automation in Construction*, 150, 104836. https://doi.org/10.1016/j.autcon.2023.104836

Makransky, G., Lilleholt, L. (2018). Emotional value of immersive VR in education. *Computers & Education*, 120, 38–52. https://doi.org/10.1007/s11423-018-9581-2

Martínez-Gutiérrez, A., Díez-González, J., Perez, H., Araújo, M. (2024). Towards Industry 5.0 through metaverse. *Robotics and CIM*, 89, 102764. https://doi.org/10.1016/j.rcim.2024.102764

Mastrolembo-Ventura, S., Castronovo, F., Nikolić, D., Ciribini, A. L. C. (2022). VR in construction education: A review. *ITcon*, 27, 705–731. https://doi.org/10.36680/j.itcon.2022.035

Meta Support. (2025). *Meta Quest 2: Tech specs*. `https://www.meta.com/us/es/quest/products/quest-2/tech-specs`

Meta. (2025). *Primeros pasos con Meta Quest 2*. `https://www.meta.com/mx/quest`

Microsoft. (2025). *Xbox wireless controller*. `https://www.xbox.com/es-mx/accessories/controllers/xbox-wireless-controller`

Mozilla. (2024). *Mozilla Hubs: Open-source WebXR virtual spaces*. `https://hubs.mozilla.com`

Olivarría González, M. C., Peraza Garzón, J. F., Valenzuela Bañuelos, A., Tostado Ramírez, M. I. (2023). Creación de un metaverso con OpenSimulator. *ReDTIS*, 7(1), 35–42. https://doi.org/10.61530/redtis.vol7.n1.2023.185.35-42

OpenSimulator. (2024). *OpenSimulator—Download*. `http://opensimulator.org/wiki/Download`

Parong, J., Mayer, R. E. (2018). Learning science in IVR. *Journal of Educational Psychology*, 110(6), 785–797. https://doi.org/10.1037/edu0000241

Pellas, N., Dengel, A., Christopoulos, A. (2021). Immersive VR in STEM: A scoping review. *IEEE TLT*, 13(1), 45–66. https://doi.org/10.1109/TLT.2020.3019405

PICO Interactive. (2022). *PICO 4: White paper*. `https://www.picoxr.com/global/pico4`

Radianti, J., Majchrzak, T. A., Fromm, J., Wohlgenannt, I. (2020). Review of IVR in higher education. *Computers & Education*, 147, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rebenitsch, L., Owen, C. (2016). Individual variation in cybersickness. *ACM SAP*, 19–23. https://doi.org/10.1145/2642918.2647394

Saenz, D. (2025). *Official OpenSim documentation*. OpenSimulator Wiki. `http://opensimulator.org/wiki/Portada`

Salatino, A., Zavattaro, C., Gammeri, R., et al. (2023). VR rehabilitation for spatial neglect: A review. *Neuroscience & Biobehavioral Reviews*, 152, 105248. https://doi.org/10.1016/j.neubiorev.2023.105248

Sosa Jiménez, E., López Martínez, J., Aguilar Vera, R. (2020). RA en educación: Revisión sistemática. *Tecnología Educativa Revista CONAIC*, 7(2), 19–30.

Stemfinity. (2025). *ClassVR Premium Headset*. Retrieved from `https://stemfinity.com/products/classvr-premium-headset-64gb-eduverse-license-bundle`

Stracke, C. M., Bothe, P., Adler, S., et al. (2025). Immersive VR in higher education: Review. *Virtual Reality*, 29, 64. https://doi.org/10.1007/s10055-025-01136-x

Sweeney, T. (2021). *The metaverse: The ultimate platform*. Epic Games. `https://www.epicgames.com/help/es-ES/`

Tlili, A., Huang, R., Shehata, B., et al. (2022). Is metaverse in education a blessing or a curse? *Smart Learning Environments*, 9(1), 24. https://doi.org/10.1186/s40561-022-00205-x

Unity Technologies. (2024). *Unity XR Interaction Toolkit*. `https://unity.com/products/xr-interaction-toolkit`

Valve Corporation. (2019). *Index VR headset documentation*. `https://www.valvesoftware.com`

Valve Corporation. (2025). *SteamVR*. `https://store.steampowered.com/app/250820/SteamVR`

Varjo. (2023). *XR-4 series datasheet*. https://www.varjo.com

Won, M., Ungu, D. A. K., Matovu, H., et al. (2023). Approaches to learning with IVR. *Computers & Education*, 195, 104701. https://doi.org/10.1016/j.compedu.2022.104701

Wonda VR. (2024). *Pricing plans*. https://wondavr.com

Wu, B., Yu, X., Gu, X. (2020). Effectiveness of IVR on learning performance. *British Journal of Educational Technology*, 51(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Zapata, J. (2024). MySQL vs PostgreSQL: Response time comparison in e-commerce. City College Los Angeles. https://doi.org/10.13140/RG.2.2.24791.69288

Zuckerberg, M. (2021). *Building the metaverse*. https://ctu.int/wp-content/uploads/2022/01/LESTER-GARCIA-FROM-META-Constr_Metaverse_policy_spectrum

Received September 30, 2025 , revised December 11, 2025, accepted January 12, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 27-49
https://doi.org/10.22364/bjmc.2026.14.1.02

Acoustic Features of Lithuanian Directives

Asta KAZLAUSKIENĖ

Vilnius University, Universiteto str. 5, LT-01513 Vilnius, Lithuania

asta.kazlauskiene@vdu.lt

ORCID 0000-0002-6135-1635

Abstract. The aim of this study is to identify the acoustic similarities and differences among directives. The data consist of specially constructed sentences read by four male actors, representing three subtypes of directive: request, command, and advice (a total of 340 tokens). The following parameters were analysed: mean F0, F0 maximum, F0 range, intonation contour, mean intensity and intensity maximum, as well as the duration of some segments. The results show that the parameters of F0 vary, differ only slightly in some cases, and do not reveal clear tendencies. The intonation contour is falling or rising–falling in the examples containing two intermediate phrases. Intensity is a reliable indicator of active directives – commands and pleas. Phrases expressing advice show the lowest intensity values. Longer duration signals requests.

Keywords: Lithuanian, directive, fundamental frequency (F0), intonation contour, intensity, duration.

1. Introduction

Directives express various degrees of encouragement or prohibition. In Lithuanian, grammatical expression of directive force may be conveyed by imperative sentences employing the imperative verb forms, as well as by declarative sentences containing verbs with a specific performative meaning. For example: *Perrašyk darbą.* 'Rewrite the paper.', *Prašyčiau perrašyti darbą.* 'I would ask you to rewrite the paper.', *Prašau perrašyti darbą.* 'I ask you to rewrite the paper.' In speech, the grammatical and lexical means of expressing directives can be supplemented by phonetic features and non-linguistic means (gestures, facial expressions, etc.).

In many languages, e.g., Romance languages (Frota and Prieto, 2015), German, and Norwegian (Cruttenden, 1997), the intonation patterns of directives resemble those of statements: the pitch either falls at the end of the phrase or remains low throughout the utterance. In some cases, however, the pitch may rise slightly, depending on the function of the directive. In Norwegian (Cruttenden, 1997) and German (Gibbon, 1998), more forceful directives, such as commands, are produced with a falling pitch, whereas milder directives, such as requests, tend to be expressed with a slightly rising pitch. The fundamental frequency (F0) range and register may help to distinguish directives from one another and from declaratives (Falé and Faria, 2007). Directives typically have a somewhat wider F0 range and a higher register. Other important properties for recognising directives include intensity and/or voice quality features (for Swedish, see

Gårding, 1998), as well as the lengthening of certain syllables (for Romance languages, see Frota and Prieto, 2015).

Relatively few studies have been devoted to the acoustic properties of directives in Lithuanian. The findings of P. Bikulčienė's (1978) research show that directives are characterised by: a) a falling intonation contour at the end of the phrase; the intonation contour is either rising–falling (typical of prohibitions, commands, and demands) or falling (typical of requests and advice) with a pitch peak in the stressed or pre-stress syllable (cf. E. Mikalauskaitė (1975, 84), who argued that directive sentences are characterised by rising intonation); b) decreasing intensity towards the end of the phrase, with the intensity peak in the pre-stress or stressed syllable; c) relatively longer duration of stressed syllables; d) variation in the F0 maximum depending on the subtype of directive. On this basis, the following descending order of F0 maximum can be established: command, advice, request.

Kundrotas (2018, 43–116) discusses Lithuanian intonation types, some of which are also typical of directives: a) for polite commands or requests, the pitch rises up to the intonation centre and then falls; b) for commands with an invitational, permissive, or instructive character, the pitch falls up to the intonation centre, the centre itself is falling–rising, and after the centre the pitch rises; c) for wishes, desires, and well-wishes there are two intonation centres; in the first centre the pitch is rising, and it is falling in the second centre; the pitch rises up to each centre and falls after them; d) for emphasised directives and commands, the pitch rises up to the intonation centre, which may be rising or rising–falling; after the centre the pitch falls.

2. Aim and Methodology

The aim of the study outlined in this article is to determine the acoustic similarities and differences between directives. This study is part of a larger project that examines the acoustic properties of Lithuanian functional phrase types. A central tenet of the project is its applied nature, as the goal encompasses not solely the identification of the properties of directives but also their description in a manner that ensures their practical application in the enhancement of speech technology tools. This means that, to facilitate the automatic recognition or synthesis of functional phrase types, it is necessary to identify and model acoustic parameters directly from the speech signal. For this reason, the studies by Lithuanian authors previously referenced offer a useful starting point for further work. However, the direct applicability of their findings is limited (Bikulčienė's due to the analysis of one-word phrases, Kundrotas's due to the absence of acoustic data).

It should be noted that in this part of the research directives are not compared with statements or questions. Consequently, the results are evaluated and interpreted only within the set of directive examples, i.e., by comparing directive phrases with one another. This approach inherently constrains the range of interpretive possibilities, as it does not employ statements as a baseline. However, this methodological choice enables the analysis to concentrate on a specific functional phrase type. Once all three types – statement, question, and directive – have been analysed, it will be possible to provide a detailed description of their similarities and differences in Lithuanian.

The classification of imperative sentences in Lithuanian linguistics is not uniform. Lithuanian researchers in syntax and phonetics (Balkevičius, 1963, 56–64; Vaitkevičiūtė, 1964; 1966; 2001, 166–179; Sirtautas and Grenda, 1988, 18–21; Oginskienė, 1994, 578–

580; Labutis, 1998, 116–119) divide imperative sentences into two groups: imperatives (active directives) and optatives (more passive directives). Bikulčienė (1978), however, divided directive phrases (not imperative sentences) into three groups: requests, commands, and advice. The present study is based on this latter classification when constructing the research material. At the same time, the analysis also considers the degree of directive force, distinguishing between more active and less active directives.

For the purposes of this study, the following sentences were constructed (IP = intonational phrase):

IP1: *Perrašyk darbą.* 'Rewrite the paper.' (pronounced as a request)
IP2: *Perrašyk darbą.* 'Rewrite the paper.' (pronounced as a command)
IP3: *Perrašyk darbą.* 'Rewrite the paper.' (pronounced as advice)
IP4: *Jis prašo* "*Perrašyk darbą*". 'He asks, "Rewrite the paper."
IP5: *Jis reikalauja* "*Perrašyk darbą*". 'He demands, "Rewrite the paper."
IP6: *Jis pataria* "*Perrašyk darbą*". 'He advises, "Rewrite the paper."
IP7: *Prašyčiau perrašyti darbą.* 'I would ask you to rewrite the paper.'
IP8: *Liepčiau perrašyti darbą.* 'I would order you to rewrite the paper.'
IP9: *Patarčiau perrašyti darbą.* 'I would advise you to rewrite the paper.'
IP10: *Prašau perrašyti darbą.* 'I ask you to rewrite the paper.'
IP11: *Maldauju perrašyti darbą.* 'I beg you to rewrite the paper.'
IP12: *Liepiu perrašyti darbą.* 'I order you to rewrite the paper.'
IP13: *Reikalauju perrašyti darbą.* 'I require you to rewrite the paper.'
IP14: *Siūlau perrašyti darbą.* 'I suggest rewriting the paper.'
IP15: *Patariu perrašyti darbą.* 'I advise you to rewrite the paper.'
IP16: *Perrašyk darbą!* 'Rewrite the paper!' (pronounced expressively, as a shout)

From a semantic-pragmatic perspective, a) the request group is represented by IP1, IP4, IP7, IP10, IP11, b) the command group by IP2, IP5, IP8, IP12, IP13, and c) the advice group by IP3, IP6, IP9, IP14, IP15. The phrase IP16, to be pronounced expressively (in a shout-like manner), may be considered a member of the command group. However, in the presentation of the results, it is discussed separately from the command group.

Interrogative sentences are also considered directives (cf. Drukteinis, 2024, 610–616). However, they are not included in this stage of the study described in this article and will be analysed together with interrogative sentences marking the function of questioning.

The sentences used in the experiment were constructed in such a way as to allow the identification not only of semantic-pragmatic factors, but also, to some extent, of the possible interplay between grammatical factors and acoustic properties. For this reason, both the performative verb and its grammatical expression vary: a) imperative, b) subjunctive and infinitive, c) present tense and infinitive (in (a) – imperative sentence; in (b) and (c) – declarative sentences). Due to these different grammatical realisations of directives and because of the focus on acoustic properties, the term *directive phrase* is used in the article instead of *imperative sentence*.

The prepared sentences were read five times each by four male actors (aged 30–33). The recordings were made in a professional sound studio; the technical recording conditions were identical for all material.

The empirical data were extracted using the speech analysis software Praat (Boersma and Weenink, 2018), and several measures (mean and maximum intensity, duration) were obtained using the ProsodyPro script (Xu 2013). For the interpretation of certain

results, a supplementary questionnaire survey, completed by three phoneticians, was also employed.

In the article the results are described not in terms of absolute acoustic parameter values (F0 measured in Hz, intensity in dB, duration in ms), but in terms of their ratios, except for intonation contours, where F0 was measured in semitones (st). This approach is intended not only to highlight the similarities and differences between phrases more apparent and to abstract as far as possible from individual speakers' vocal properties and accidental non-linguistic factors, but also to address the applied aspect of the study mentioned above.

Ratios for the mean F0, mean intensity, and duration were calculated as follows: the mean value of the parameter under investigation was determined for the analysis unit (phrase or part of phrase), the median of the mean values of that parameter for the entire sample for a given speaker was then computed, and the mean value for each phrase (or segment) was divided by this median. For the maximum, the maximum values were divided by the median of the sample means (not by the median of the maximum). Intensity ratios were calculated using the formula $10^{(x-y)/10}$, where *x* is the intensity of the unit under investigation and *y* is the median of the sample.

F0 range[1] was calculated as F0 maximum minus F0 minimum, and range ratios were obtained by dividing this value by the median of the speaker's ranges. The centre of the range, which allows one to evaluate the height of the F0 register, was calculated as F0min + (F0max − F0min)/2, and its ratios were computed by dividing this value by the median of the mean F0 values.

The register of the intonation contour is defined as the interval between the highest and lowest F0 values. It was calculated based on F0 measurements obtained from the target syllables used to construct the contour. The excursion size of the intonation contour was calculated as the difference between the highest and the lowest F0 values along the contour.

The data were statistically evaluated using the *t-test* (significance level 0.05; unless otherwise indicated, the critical *t*-value is 2.02). A cluster analysis was conducted using IBM SPSS (IBM Corp. 2020).

3. Results and Discussion

3.1. Fundamental Frequency and Intonation contour

The results for phrase F0 are not straightforward: among those with a mean higher than the overall median of mean F0 of the speaker (ratio > 1), there are examples from all semantic-pragmatic groups under investigation – advice, command, and request (see Table 1). Nevertheless, some tendencies can be observed: the participants of the study were generally not inclined to produce advice phrases with an F0 higher than the overall median (0.93–0.99), except for IP14 with the performative verb *siūlau* 'I suggest'. Command phrases, except for IP5, tended to be produced with higher F0 values than most advice phrases; however, the interval of ratios is quite wide – from 0.86 to 1.05. Mean F0 values for request phrases also vary; there are examples both above and below the overall median (0.93–1.03). It is noteworthy that the pleading phrase (IP11) has a

[1] On the terms *pitch register* and *pitch range*, see Féry (2017, 21) and Rietveld and Vermillion (2004).

comparatively high mean F0, like that of command phrases, which represent active directives.

Table 1. Ratios of F0 parameters[2]

Phrase	Mean	Maximum	Range	Range centre
IP16	1.15	1.43	1.18	1.11
IP14	1.08	1.59	1.42	1.11
IP8	1.05	1.53	1.39	1.11
IP13	1.04	1.28	0.98	0.97
IP11	1.03	1.28	0.97	0.97
IP12	1.02	1.37	1.06	1.05
IP1	1.01	1.34	0.99	1.02
IP2	1.00	1.25	0.94	0.99
IP10	0.99	1.27	0.90	0.96
IP9	0.99	1.29	0.90	1.01
IP6	0.96	1.25	0.97	1.01
IP7	0.96	1.29	1.01	1.01
IP3	0.94	1.42	1.12	1.04
IP15	0.93	1.23	0.83	0.95
IP4	0.93	1.13	0.71	0.91
IP5	0.86	1.06	0.65	0.86

When the ratios of mean F0 of phrases are summarized by directive group, the group mean F0 can be said not to differ statistically significant[3]: IPcomm (1.00) / IPadv (1.00) > IPreq (0.99). Due to phrases that deviate from the general pattern (IP14, IP5, and IP11), the summarized results should be interpreted with caution.

The results for F0 maximum do not fully correspond to the means F0: IPadv (1.36) > IPcomm (1.28) / IPreq (1.28). The differences between IPcomm and IPadv (t = 2.44) and between IPreq and IPadv (t = 3.33, in both cases with a critical t-value = 1.97) are statistically significant.

Among the phrases with the highest maximum is IP3, which is read as advice and has one of the lowest mean F0 values. In addition, F0 values for this phrase show very large dispersion, reflecting different realisation strategies: starting the phrase with a higher or somewhat lower F0 (see Figures 1 and 2). It is precisely this phrase, together with the already mentioned IP14 with *siūlau* 'I suggest' that produced the high overall value for advice.

Phrases IP10–IP15 are declarative sentences containing performative verbs that convey directives of varying degrees of force: *prašau* 'I ask' (IP10) and *maldauju* 'I beg'

[2] In this and in Table 2, the phrases are ordered according to the ratios of the means, from the largest to the smallest. The phrases are arranged in the same order in the box plots in Figures 1, 2, 7, and 8.

[3] Here and in what follows, unless indicated otherwise, the value given in brackets is the median of the ratios; the symbol ">" is used when the values for the compared phrases differ (although the difference is not necessarily statistically significant), and "/" when they are equal. IPcomm is used as a cover label for command phrases, IPadv for advice phrases, and IPreq for request phrases.

(IP11), *reikalauju* 'I demand' (IP13) and *liepiu* 'I order' (IP12), *patariu* 'I advise' (IP15) and *siūlau* 'I suggest' (IP14). According to the assessments of phoneticians who participated in the questionnaire survey, the second verb in each pair is semantically stronger in terms of encouragement. It could have been expected that not only command phrases in general, but also those containing these semantically more active performative verbs would support Bikulčienė's (1978) conclusion that active directives are characterised by a higher F0 maximum. However, this is not the case: among the compared pairs, only IP14 exhibits a higher maximum than IP15; the maximum of IP11 and IP10 are very similar, whereas IP13 has a higher maximum than IP12. These findings indicate that F0 is neither the sole, nor perhaps necessarily the primary indicator of directive force.

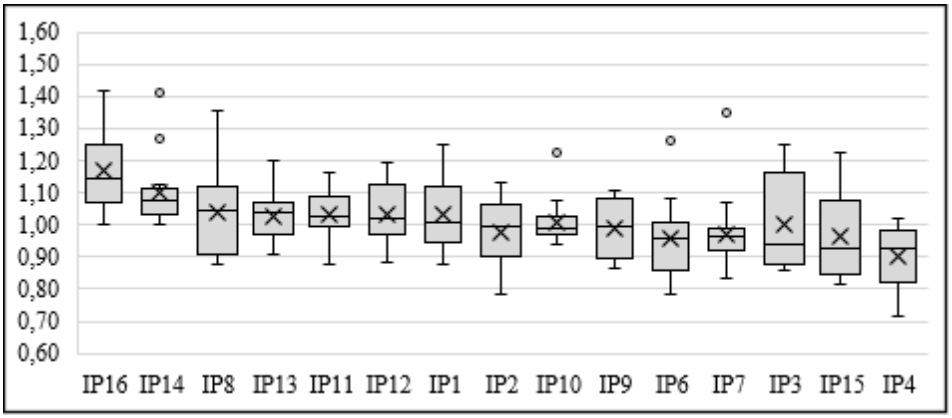


Figure 1. Ratios of mean F0

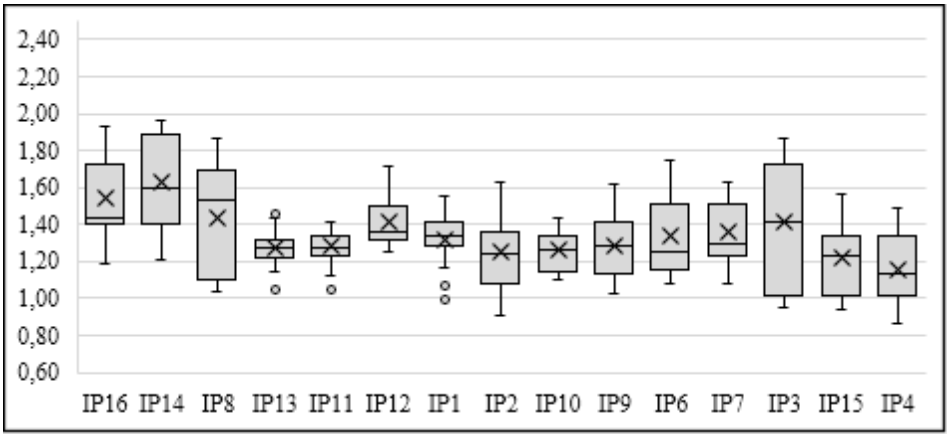


Figure 2. Ratios of F0 maximum

The high maximum of some advice phrases contributed to the widest F0 range among the compared groups (1.06). The differences between IPadv and IPcomm, and between IPadv and IPreq are statistically significant (t = 2.11 and t = 2.95, respectively; critical t-value = 1.97). Request phrases have the narrowest range (0.93), whereas command phrases (0.97) occupy an intermediate position between advice and request.

The overall ratios of F0 range centres for command and request phrases are identical (0.98) and lower than for advice (1.02). The differences between IPadv and IPcomm, and between IPadv and IPreq are statistically significant (t = 2.4 and t = 2.96, critical t-value = 1.97). Thus, in the present material advice phrases are characterised not only by the widest range, but also by a somewhat higher register. A close examination of phrases with high F0 maximum shows that phrases with an F0 maximum at least 1.4 times higher than the mean were produced in a higher register. In this way the participants tended to produce not only the expressive command IP16, but also advice with *siūlau* 'I suggest' (IP14), a command with *liepčiau* 'I would order' (IP8), as well as some individual phrases. However, it cannot be claimed that these phrases were produced with strongly emphasised narrow focus, since in Lithuanian narrow focus is signalled by a complex set of features: not only F0, but also higher intensity and longer duration (Kazlauskienė and Dereškevičiūtė, 2022; Kazlauskienė et al. 2023, 37–86).

Four phrases with the same phonemical structure, *Perrašyk darbą* 'Rewrite the paper', were pronounced differently: expressively (IP16), as a request (IP1), as a command (IP2), and as advice (IP3). All F0 parameters of the expressively produced phrase are comparatively high and therefore stand out from the other three realisations. For the non-expressively produced phrases, the ordering of F0 parameters varies: mean F0 – IP1 > IP2 > IP3; F0 maximum and F0 range – IP3 > IP1 > IP2. Although none of these differences is statistically significant, several observations can still be made. As previously mentioned, IP3 has a lower mean F0 than IP1 and IP2 (the latter two are similar), but it exhibits the highest maximum and range.

This suggests that the acoustic realisation of advice lacks consistency and shows substantial variability; speakers may perceive and produce advice differently, particularly with respect to its degree of forcefulness. At the same time, these findings support the cautious conclusion that the three directive realisations of a phrase with identical structure are unlikely to be distinguishable based on F0 alone[4].

In the material three pairs of phrases contain the same performative verb in the subjunctive and present tense. In one pair (IP10 > IP7, *prašau* 'I ask' vs. *prašyčiau* 'I would ask') the phrase with the present tense shows a higher mean F0, although the F0 maximum is higher in the sentence with the subjunctive form of the verb. In the remaining pairs, both the mean and the maximum F0 are higher in phrases with the subjunctive: IP8 > IP12 and IP9 > IP15 (*liepčiau* 'I would order' vs. *liepiu* 'I order', *patarčiau* 'I would advise' vs. *patariu* 'I advise'). However, neither the individual differences within the pairs nor the overall difference between the two groups is statistically significant. Therefore, only a tentative observation can be made. Directives expressed through sentences containing performative verbs in the subjunctive may be produced with higher F0 than directives with performative verbs in the present tense. If

[4] That these phrases are not easily recognisable in the material is further supported by the results of a brief survey conducted among phoneticians: a) only half of the request phrases were identified as requests, while the others were classified as commands or advice; b) two thirds of the command phrases were correctly identified as commands, while the rest were perceived as requests; and c) nearly three fifths of the advice phrases were identified as advice, while the others were perceived as requests.

higher F0 is taken as an indicator of a more active directive (cf. most command phrases have higher F0 parameters), phrases with performatives in the subjunctive may be selected to express a stronger directive force. Nevertheless, this insight is not supported by the phoneticians who participated in the survey. In their judgement, more active directives are those expressed with present tense performative verbs, although they also pointed to certain contextual factors, such as the influence of social roles on directive realisation: speakers in positions of authority may be more likely to choose the present tense.

Within each directive group, it is possible to compare two additional types of realisations: those without a performative verb and those containing a performative verb in another clause of a compound sentence. The latter constructions are close to direct speech. All speakers produced them as two intermediate phrases, even when no clear pause separated the two. The results of this comparison do not reveal any consistent pattern, except that in both the request and command groups the lowest F0 values occurred in the examples realised with two intermediate phrases:

a) IP1 > IP10 > IP7 > IP4 (read as a request, *prašau* 'I ask', *prašyčiau* 'I would ask', *prašo* 'He asks'),

b) IP8 > IP12 > IP2 > IP5 (*liepčiau* 'I would order', *liepiu* 'I order', read as a command, *reikalauja* 'He demands'),

c) IP9 > IP6 > IP3 > IP15 (*pataršiau* 'I would advise', *pataria* 'He advises', read as advice, *patariu* 'I advise').

This suggests a cautious hypothesis that phrases with the performative verb in the third person are less expressive in terms of F0. In these cases, the directive is indirect – the speaker does not issue the directive personally but reports someone else's directive. However, only a few of the F0 differences between the compared examples are statistically significant: IP4 and IP1 (t = 4.00), IP4 and IP7 (t = 2.08), IP4 and IP10 (t = 4.02), IP8 and IP5 (t = 4.67), IP5 and IP12 (t = 5.26).

The studies reviewed in the introduction suggest that the intonation contour, particularly the F0 height at the end of the phrase, may vary depending on the type of directive. To evaluate whether such assumptions hold for Lithuanian, F0 (measured in semitones relative to 1 Hz) was extracted for specific syllables, and median values were calculated:

a) for IP1–IP3 and IP16, F0 was measured in the nucleus of the first syllable, the final stressed syllable, and the final syllable of the phrase (***P**e**r**rašyk d**ar**bą*** 'Rewrite the paper'; target syllables are indicated in bold).

b) for IP4–IP15, F0 measurements were taken from the stressed syllable of the performative verb and from the corresponding syllables in ***p**e**r**rašyk d**ar**bą*** 'rewrite the paper'.

The results for phrases IP1–IP3 and IP16 indicate (see Figure 3) that they are characterised by a falling contour. IP16 is produced in a slightly higher register than the other phrases (77–89 st, compared with overlapping registers for the others: 75–87 st (IP1), 76–87 st (IP2), 74–85 st (IP3)), yet its F0 excursion is identical to that of the request and advice phrases – 12 st; only the command phrase IP2 exhibits a smaller F0 excursion (9 st).

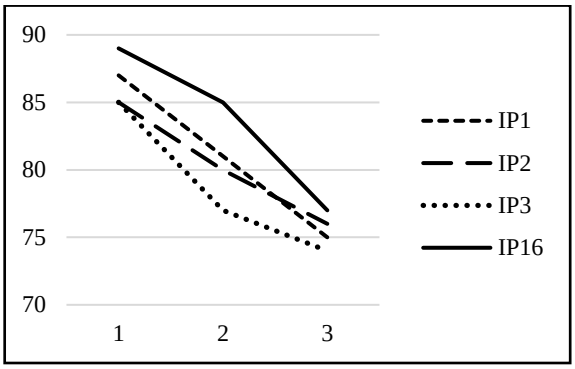


Figure 3. Intonation contour in phrases IP1–IP3 and IP16[5]

For IP4–IP6, which consist of two intermediate phrases and whose second phrase corresponds to that of IP1–IP3, the contour is rising in the first intermediate phrase and falling in the second[6] (see Figure 4). Comparing the contour of the second intermediate phrase (*perrašyk darbą* 'rewrite the paper') with IP1–IP3, it can be observed that the curves for advice and request are similar, differing only in F0 excursion: IP4 and IP6 have excursions more than twice as small – only 5 st. The contour of the command phrase IP5 is not as steep as that of IP2, although both have smaller F0 excursion than the other phrases in their respective groups (4 st for IP5). The request phrase IP4 is pronounced in a slightly higher register (78–86 st), cf. the registers of the advice phrase IP6 (77–82 st) and the command phrase IP5 (78–83 st).

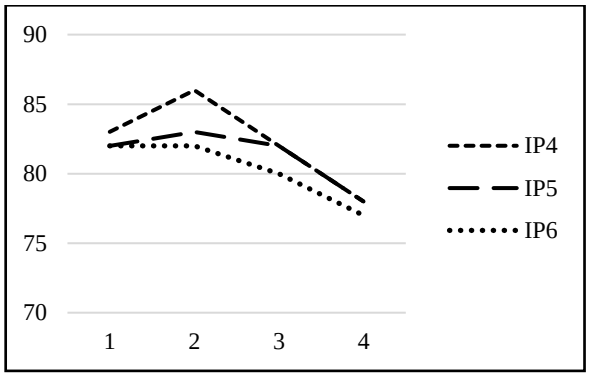


Figure 4. Intonation contour in phrases IP4–IP6

[5] In this figure, the digits on the x-axis indicate the syllables of *P**e**rrašyk d**ar**bą* 'Rewrite the paper' in which the F0 was measured (target syllables are indicated in bold). In Figures 4–6, 1 corresponds to the stressed syllable of the performative verb and 2–4 to the syllables of *p**e**rrašyti d**ar**bą*.

[6] The F0 of the first word *jis* ('he') was also measured, but the values did not differ from those of the performative verb and were therefore excluded from further analysis.

The contours of IP7–IP9, which include performative verbs in the subjunctive, are similar to those shown in Figure 3, differing primarily in excursion: IP8 and IP9 exhibit somewhat steeper excursions (11 and 13 st, respectively, compared with 9 st for IP2 and 11 st for IP3), whereas IP7 has a notably smaller excursion of 7 st (compared with 12 st for IP1). The registers of these phrases vary: 78–85 st for IP7, 76–87 st for IP8, and 73–86 st for IP9.

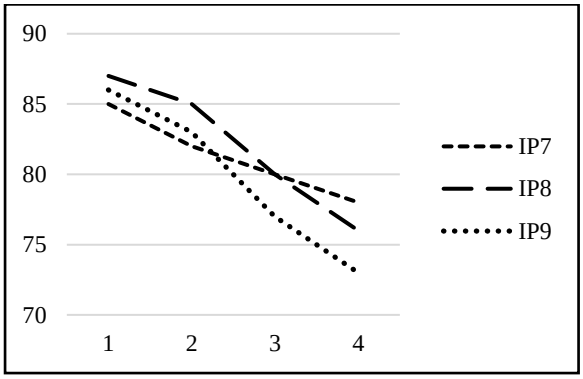


Figure 5. Intonation contour in phrases IP7–IP9

The contours of IP11–IP15 (see Figure 6) differ little from those of the phrases shown in Figure 5. Only IP10 shows a rising–falling contour. The F0 excursions in these examples range from 9 to 10 st, except for the advice phrase IP14, which has an excursion of 12 st. The registers of the command phrases IP12 and IP13 are identical (77–87 st). Whereas the register of the advice phrases varies considerably: 76–88 st for IP14 and 75–85 st for IP15. The register of the request phrase IP11 is 76–86 st.

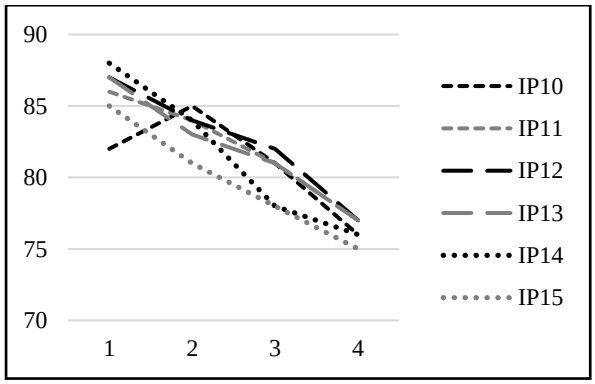


Figure 6. Intonation contour in phrases IP10–IP15

The expressively produced phrase begins with the highest F0 (89 st). Among the non-expressively produced phrases without a performative verb, the request phrases have the highest initial F0 (87 st), while the command and advice phrases start at the same level (85 st). For the examples containing two intermediate phrases, the initial F0 of the request is also somewhat higher, both in the first intermediate phrase (IP4 – 83 st, compared with 82 st for IP5 and IP6) and in the second (IP4 – 86 st, compared with 83 st for IP5 and 82 st for IP6).

Among the phrases with performatives in the subjunctive, the highest initial F0 is found for the command (87 st for IP8, 86 st for IP9 and 85 st for IP7). Commands with present tense performatives (IP12 and IP13) also show a high initial F0 (87 st). However, the highest initial F0 is observed in IP14 (88 st, cf. 85 st for the other advice phrase IP15). For the request phrases of this structural group, the initial F0 is not uniform: IP11 – 86 st, IP10 – 82 st.

Across all structurally different groups, the final F0 is lowest in advice phrases (values in st are given in brackets): a) IP2 (76) > IP1 (75) > IP3 (74), b) IP4 / IP5 (78) > IP6 (77), c) IP7 (78) > IP8 (76) > IP9 (73), d) IP11 / IP12 / IP13 (77) > IP10 / IP14 (76) > IP15 (75).

In a few cases, participants produced a slightly rising intonation contour at the very end of the phrase (approximately 1–1.5 st). However, it is not possible to conclude that this occurred more frequently in any particular directive type; such cases appeared across all groups and are likely related not to directive force, but to non-final (list-like) intonation.

When comparing F0 in the stressed syllables of the performative verbs one tendency can be observed: speakers tended to emphasize present tense verbs in the first person more than those in the third person:

a) *maldauju* ‘I beg’ (86) > *prašyčiau* ‘I would ask’ (85) > *prašo* ‘He asks’ (83) > *prašau* ‘I ask’ (82),

b) *liepiu* ‘I order’ / *reikalauju* ‘I require’ / *liepčiau* ‘I would order’ (all 87) > *reikalauja* ‘He demands’ (82),

c) *siūlau* ‘I suggest’ (88) > *patarčiau* ‘I would suggest’ (86) > *patariu* ‘I advise’ (85) > *pataria* ‘He advises’ (82).

As can be seen, *prašo* (‘He asks’) is an exception, as its F0 is higher than that of *prašau* (‘I ask’); the latter was pronounced most likely as an interjection.

To summarise the F0 results: command phrases tend to be produced with higher F0. However, the main reason for higher F0 is not the type of directive, but rather expressiveness and the height of the phrase-initial F0. When a phrase begins at a higher F0 and the performative verb is somewhat more emphasized, all F0 parameters, especially the maximum and range, are high regardless of directive type. For this reason, there is considerable variability in phrase-initial F0, since the performative verb occurs in phrase-initial position in the material. Thus, phrases expressing a request, a command, or advice may all start with high F0. Advice phrases have the lowest final F0, whereas command and request phrases are more variable, and no consistent pattern emerges.

The intonation contour in the material is predominantly falling; a rising–falling contour was observed in examples with two intermediate phrases, as well as in the phrase *Prašau perrašyti darbą.* ‘I ask you to rewrite the paper.’ These results are fully consistent with the findings of the foreign studies cited at the beginning of the article (Cruttenden, 1997; Simard, 2013; Frota and Prieto, 2015). However, they only partially support Bikulčienė’s (1978) observation that active directives are characterised by a

rising–falling intonation contour, whereas less active ones, such as requests, show a falling contour.

3.2. Intensity

When the ratios of mean intensity are calculated and the phrases are ordered accordingly, those belonging to the command group cluster at the top of the list (see Table 2). Among the phrases with the highest intensity is also a request containing the performative verb *maldauju* 'I beg'. In general, command phrases and this pleading phrase exhibit the largest variability in intensity values (see Figures 7 and 8), indicating that the participants employed quite different strategies for expressing commands: some produced them with substantially high intensity, while others deviated only slightly from their typical loudness.

Table 2. Ratios of intensity

Phrase	Mean	Maximum	Phrase	Mean	Maximum
IP16	2.91	14.30	IP4	0.85	3.96
IP11	2.28	11.13	IP6	0.78	4.35
IP13	2.23	11.38	IP10	0.75	2.81
IP2	1.85	7.91	IP9	0.64	3.38
IP12	1.58	8.47	IP14	0.57	2.78
IP5	1.49	6.17	IP1	0.48	1.77
IP8	1.34	8.21	IP15	0.48	2.65
IP7	1.06	5.50	IP3	0.39	2.03

Participants tended to produce advice with lower than mean intensity, with intensity ratios ranging narrowly from 0.39 to 0.78 (cf. command phrases, where the range, excluding the expressive phrase, spans 1.34 to 2.23).

Except for the phrase containing *maldauju* 'I beg', the intensity of request phrases is below mean intensity. Only the request with the performative verb *prašyčiau* 'I would ask' exhibits intensity close to mean. Overall, the intensity of request phrases is highly variable: they appear among both the high- and low- intensity phrases, with values ranging from 0.48 to 2.28 (for the pleading phrase).

These tendencies are reflected in the summarized intensity results for the directive groups. Commands (excluding the expressive example) exhibit the highest intensity, averaging roughly twice that of requests and three times that of advice: IPcomm (1.80) > IPreq (0.86) > IPadv (0.58). A similar pattern is observed for intensity maximum: IPcomm (8.26) > IPreq (3.96) > IPadv (3.25). All differences in mean and maximum intensity are statistically significant (critical *t*-value = 1.97): means – IPcomm and IPreq (t = 5.17), IPcomm and IPadv (t = 10.33), IPreq and IPadv (t = 4.96); maximum – IPcomm and IPreq (t = 6.50), IPcomm and IPadv (t = 8.72), IPreq and IPadv (t = 2.45).

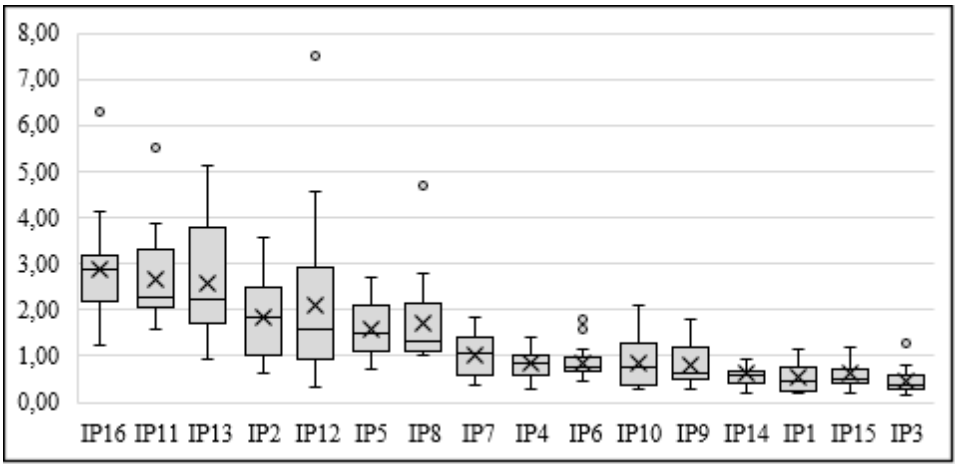


Figure 7. Ratios of mean intensity

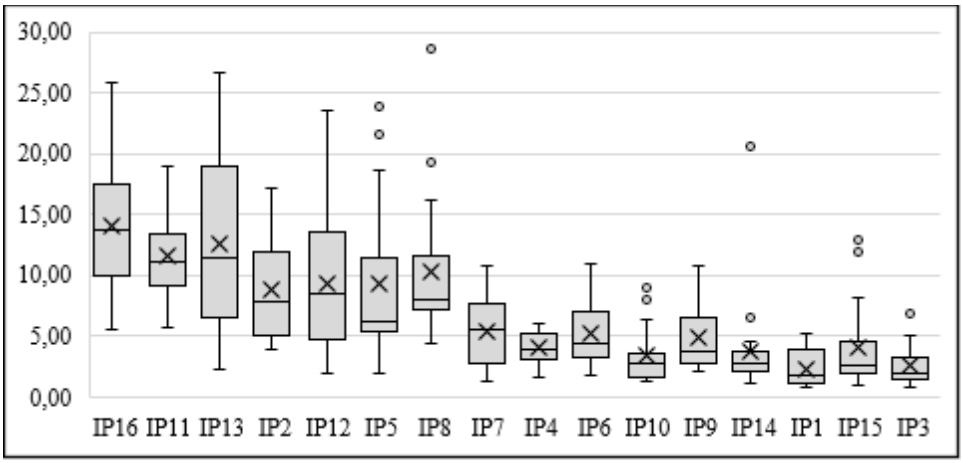


Figure 8. Ratios of maximum intensity

Phrases with the same structure, *Perrašyk darbą* 'Rewrite the paper', without performative verbs can be ordered by decreasing intensity as follows: IP16 > IP2 > IP1 > IP3 (expressive, command, request, advice). The expressive phrase has the highest intensity; and its differences from the other phrases are statistically significant. Intensity differences among the remaining phrases are smaller. However, the differences between the command and request (t = 5.95) and between the command and advice (t = 6.32) are statistically significant. These results indicate that intensity alone is unlikely to distinguish between the request and advice readings of phrases with the same structure.

When comparing phrases with the same performative verb in the present tense and subjunctive, it can be observed that in two pairs the phrases with the subjunctive are pronounced with higher intensity (IP7 > IP10, *prašyčiau* 'I would ask' and *prašau* 'I ask', IP9 > IP15, *patarčiau* 'I would advise' and *patariu* 'I advise'), whereas in one pair (IP12 > IP8, *liepiu* 'I order' and *liepčiau* 'I would order') the phrase with the present tense is more intense; however, none of these differences are statistically significant. Thus, one can speak only of a tendency for requests and advice expressed by performative verbs in the subjunctive to be produced with higher intensity than those expressed in the present tense, potentially softening the directive. Conversely, active directives – commands – are more intensely pronounced in phrases with present tense verbs and are not softened by the subjunctive. This observation should be tested on a larger and more diverse dataset.

The results of comparing fourfold different grammatical realisations of the same directive do not reveal a consistent pattern:

a) IP2 > IP12 > IP5 > IP8 (read as a command, *liepiu* 'I order', *reikalauja* 'He demands', *liepčiau* 'I would order'),

b) IP7 > IP4 > IP10 > IP1 (*prašyčiau* 'I would ask', *prašo* 'He asks', *prašau* 'I ask', read as a request),

c) IP6 > IP9 > IP15 > IP3 (*pataria* 'He advises', *patarčiau*'I would advise', *patariu*, 'I advise', read as advice).

Only the differences between IP7 and IP1 (t = 3.99) and between IP6 and IP3 (t = 4.19) are statistically significant. As can be seen, in one case the phrases without performative verbs have the highest intensity (a), in another, phrases with the subjunctive show the highest intensity (b), and in a third, phrases in the present tense display the highest intensity (c). A cautious hypothesis may be proposed: when a command is expressed without a performative verb, intensity becomes a crucial feature, and the phrase exhibits high intensity, closely resembling a prototypical command; performative verbs then serve to attenuate the directive. In request and advice phrases without performative verbs, intensity is low, and performative verbs function to strengthen the directive. In such cases, acoustic properties play a complementary role.

To summarise the intensity results: mean and maximum intensity serve as a reliable indicator for recognising not only expressively produced phrases but also for command phrases in general, that is, active directives. This conclusion is further supported by the relatively high intensity values for pleading, which also falls within the category of active directives. Moreover, it aligns with the finding that the intensity values for *Perrašyk darbą* 'Rewrite the paper' differ only slightly when produced as advice or as a request, whereas the command reading shows considerably higher intensity.

The directive can also be reinforced by the performative verb (through its semantics), especially when it is emphasized. In this study, this effect is observed for advice expressed with *siūlau* 'I suggest', particularly when the verb bears narrow focus.

A comparison of F0 and intensity maximum shows that the phrases with the highest F0 maximum do not generally coincide with those exhibiting the highest intensity maximum except for IP16 and IP8. The highest F0 maximum are observed in IP14 (1.59), IP8 (1.53), IP16 (1.43), and IP3 (1.42), whereas the highest intensity maximum occurs in IP16 (14.30), IP13 (11.38), IP11 (11.13), IP12 (8.47), and IP8 (1.21). If the phrases with the highest F0 maximum also had the highest intensity maximum, one could argue that these phrases contained very prominent narrow focus, since as noted, this kind of focus in Lithuanian is characterised by both features. The results of the present study suggest a cautious conclusion: the F0 maximum largely reflects the

speakers' tendency to begin phrases at a higher F0, whereas intensity is more closely related to the force of the directive. At the same time, these findings highlight an important methodological challenge in analysing the intonation contours of functional phrase types: how to minimise the influence of other linguistic phenomena, if this is at all possible.

3.3. Duration of Selected Units

Due to differences in composition, it is not methodologically appropriate to compare the duration of all phrases and words. However, the material was designed in such a way that the duration of a constant part of the phrase could be reliably compared. These constant parts are the sequences *perrašyk darbą* 'rewrite the paper' and *perrašyti darbą* 'to rewrite the paper'. The analysis considered not only the overall duration ratios of these units but also which segment type – vowels or consonants – is more sensitive to duration changes. For this purpose, the total duration of syllable nuclei and the total duration of syllable consonants in the word were calculated. For example, in *perrašyk darbą* 'rewrite the paper' the sum of nuclei durations is 674 ms, whereas the sum of consonants is 456 ms. Median and ratio of duration were calculated within the respective datasets (*perrašyk darbą* or *perrašyti darbą*).

Phrases containing *perrašyk darbą* 'rewrite the paper' are of two structural types: a) without a performative verb, b) with a performative verb. The first group can be ordered by decreasing duration as follows: IP1 (1.08) > IP16 (1.03) > IP3 (0.94) > IP2 (0.93). The differences between IP1 and IP2 (t = 2.08), IP1 and IP3 (t = 5.89), and between IP3 and IP16 (t = 2.72) are statistically significant. Thus, the longest phrases are those expressing a request and the expressive command (the latter showing a large dispersion in duration values; see Figure 9), whereas advice and command phrases are shorter. Since these phrases share the same phonemic structure, the number of segments cannot account for the observed duration differences.

In the second group, the greatest duration is again found in the request phrase with its duration closely matching that of the command phrase, whereas the advice phrase is noticeably shorter: IP4 (1.01) > IP5 (1.00) > IP6 (0.94). The differences between IP4 and IP6 (t = 4.71) and between IP5 and IP6 (t = 2.99) are statistically significant.

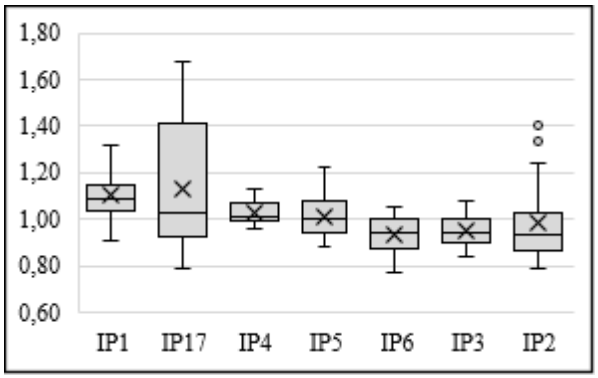


Figure 9. Duration ratios for *perrašyk darbą* 'rewrite the paper'

Based on the duration ratios of syllable nuclei (V), the phrases without performative verbs can be ordered by decreasing duration as follows: IP1-V (1.15) > IP16-V (1.08) > IP3-V (0.92) > IP2-V (0.91). The differences between IP1-V and IP2-V (t = 4.25), IP1-V and IP3-V (t = 7.15), IP3-V and IP16-V (t = 4.02), and between IP2-V and IP16-V (t = 2.85) are statistically significant. For phrases with performative verbs, the ordering is: IP4-V (1.04) > IP5-V (0.98) > IP6-V (0.89), with statistically significant differences between IP4-V and IP6-V (t = 5.47) and between IP5-V and IP6-V (t = 3.61). In both structural groups, the ordering by syllable nucleus duration corresponds to that by the total duration of the analysed part of the phrase.

A similar pattern is observed when evaluating the duration ratios of consonants (C). For phrases without performative verbs, the decreasing order is: IP1-C (1.03) > IP3-C (1.00) > IP16-C (0.97) > IP2-C (0.95); for phrases with performative verbs, the order is: IP4-C (1.01) > IP5-C (1.00) > IP6-C (0.97). The differences in both groups are small and not statistically significant.

Comparing the ratio of duration orderings for individual sounds and for *perrašyk darbą* 'rewrite the paper' shows a general tendency: the highest duration ratios are observed in request phrases (see Figure 10). However, it remains unclear whether the observed changes in phrase (and word) duration are determined by vowels or by consonants.

The ratios for syllable nuclei and consonants indicate that the difference between the largest and smallest ratio of syllable nucleus (IP1-V and IP6-V) is 0.26, whereas for consonants (IP1-C and IP2-C) it is only 0.08. This suggests that the duration of syllable nuclei varies more and may play a more significant role in the realisation of directive phrases. This pattern is also reflected in the nucleus–consonant proportions: in expressively produced and request phrases, syllable nuclei occupy a slightly larger proportion of word duration than in other directive types: IP1 – 59%, IP16 – 58%, IP4 – 57%, IP2 and IP5 – 56%, IP3 – 55%, IP6 – 54%.

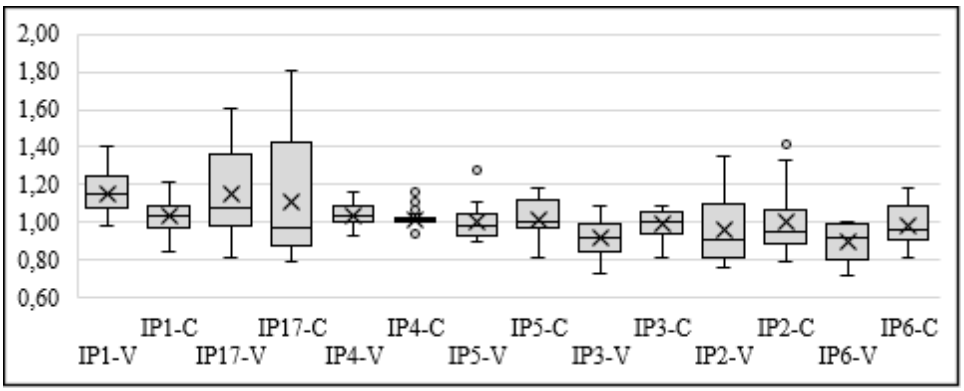


Figure 10. Duration ratios for sounds in *perrašyk darbą* 'rewrite the paper'

Phrases containing *perrašyti darbą* 'to rewrite the paper' are also of two structural types: a) with performative verbs in the subjunctive, and b) with performative verbs in the present tense. The results of duration of *perrašyti darbą* are as follows (see Figure 11): a) IP7 (1.02) > IP8 (0.96) / IP9 (0.96), b) IP10 (1.12) / IP11 (1.12) > IP12 (1.04) / IP13 (1.04) > IP15 (0.99) > IP14 (0.95). In both structural groups, the longest examples are observed in request phrases.

In the first group, the duration of *perrašyti darbą* 'to rewrite the paper' in advice and command phrases are identical, whereas in the request phrase, the sequence is significantly longer than in advice and command (IP7 and IP8 t = 2.69; IP7 and IP9 t = 2.26).

In the second group, *perrašyti darbą* 'to rewrite the paper' in advice phrases are the shortest, while those in commands occupy an intermediate position. It should also be noted that in this group, the durations of the analysed part of the request and command do not differ between phrases with different performatives, cf. IP10 and IP11 (*prašau* 'I ask' and *maldauju* 'I beg'), IP12 and IP13 (*liepiu* 'I order' and *reikalauju* 'I require'). Statistically significant differences are found between: a) IP10 and IP12 (t = 3.10), IP13 (t = 2.53), IP14 (t = 5.46), IP15 (t = 4.67); b) IP11 and IP12 (t = 2.84), IP13 (t = 2.46), IP14 (t = 4.74), IP15 (t = 4.18); c) IP12 and IP14 (t = 4.42), IP15 (t = 2.89); d) IP13 and IP14 (t = 3.54), IP15 (t = 2.57).

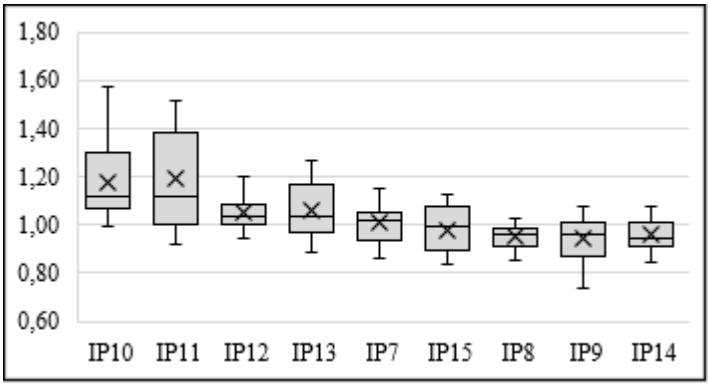


Figure 11. Duration ratios for *perrašyti darbą* 'to rewrite the paper'

The ordering of phrases by the duration of individual sounds coincides with the ordering by the duration of the entire sequence. For phrases with performative verbs in the subjunctive, the decreasing order of the ratio of syllable nucleus duration is as follows (see Figure 12): IP7-V (1.00) > IP8-V (0.94) > IP9-V (0.90). The nuclei in the request phrase are significantly longer than in advice and command (IP7-V and IP9-V, t = 3.64, IP7-V and IP8-V, t = 3.07). Consonant durations in these phrases differ slightly (IP7-C – 1.00, IP8-C and IP9-C – 0.99), and these differences are not statistically significant.

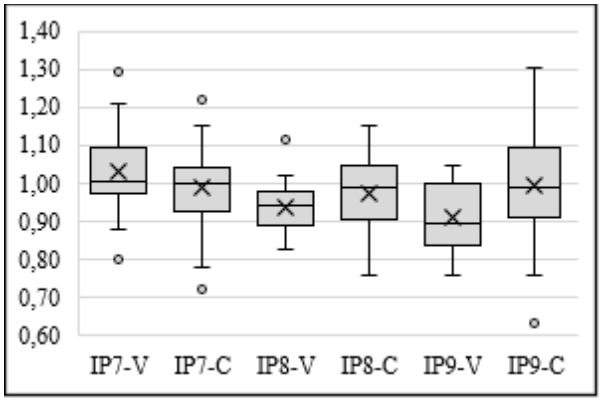


Figure 12. Duration ratios for sounds in *perrašyti darbą* 'to rewrite the paper' (1)

For phrases with performative verbs in the present tense, the decreasing order of syllable nucleus duration is as follows (see Figure 13): IP11-V (1.23) > IP10-V (1.11) > IP13-V (1.07) > IP12-V (1.03) > IP15-V (0.99) > IP14-V (0.98). Statistically significant differences are found between: a) IP10-V and IP12-V (t = 2.14), IP14-V (t = 6.92), IP15-V (t = 3.66); b) IP11-V and IP12-V (t = 2.82), IP14-V (t = 4.11), IP15-V (t = 3.94); c) IP12-V and IP14-V (t = 2.06); d) IP13-V and IP14-V (t = 2.75), IP15-V (t = 2.54). The ordering of phrases by consonant duration is very similar: IP10-C (1.12) > IP11-C (1.11) > IP12-C (1.05) > IP13-C (1.01) > IP14-C (0.97) > IP15-C (0.95). Almost all differences are statistically significant: a) IP10-C and IP12-C (t = 3.33), IP13-C (t = 4.42), IP14-C (t = 5.16), IP15-C (t = 4.78); b) IP11-C and IP12-C (t = 2.22), IP13-C (t = 3.34), IP14-C (t = 4.08), IP15-C (t = 3.77); c) IP12-C and IP14-C (t = 2.76), IP15-C (t = 2.31).

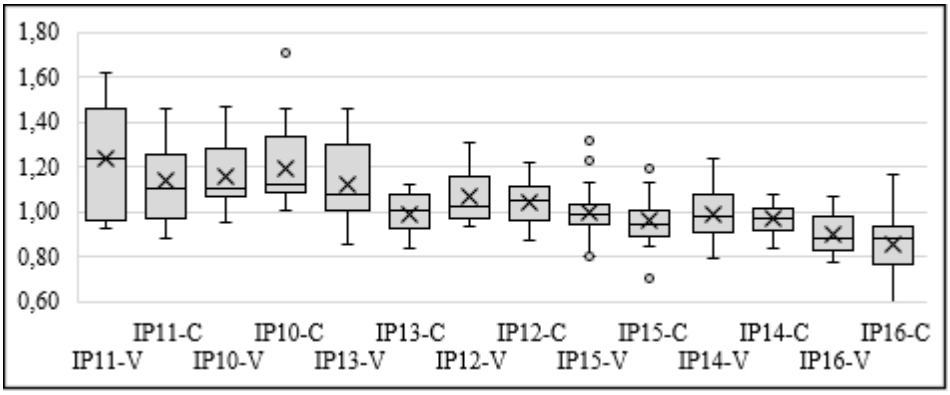


Figure 13. Duration ratios for sounds in *perrašyti darbą* 'to rewrite the paper' (2)

The proportions between syllable nuclei and consonants vary across phrases, with nuclei accounting for 54% to 59% of word duration. In the *perrašyti darbą* 'to rewrite

the paper' dataset, no clear pattern emerges when considering the proportion of nucleus duration. It may appear that syllable nuclei occupy a larger proportion of the total duration in phrases with longer verbs, e.g., trisyllabic verbs: IP13 (59%), IP11 (58%), IP7 (57%); compared with disyllabic verbs: IP12, IP14 (both 56%), IP10, IP8 (both 55%). However, phrases with trisyllabic verbs IP15 (56%) and IP9 (54%) show only a slight predominance of the nucleus.

To summarise the duration results, request phrases are consistently the longest, with all sounds lengthened, although syllable nuclei are extended slightly more. This is a reliable distinguishing feature of request phrases. Command phrases without performative verbs are the shortest, whereas those with performatives occupy an intermediate position between request and advice phrases.

3.4. Comprehensive Summary of Acoustic Features

A cluster analysis was conducted on all phrase tokens based on the relative values of seven parameters. As the results are described in terms of three semantic-pragmatic groups (request, command, and advice), the cluster analysis was set to group the phrases into three clusters. The expressively pronounced phrase was excluded from this analysis.

As shown by the summary of cluster centres and the distribution of phrases (Tables 3 and 4), the most important parameters distinguishing the clusters are intensity and duration, whereas F0 parameters show less variation.

Cluster 1 is characterised by longer duration, moderately higher F0 and range centre, and medium intensity. This cluster accounts for 28% of the data and includes many command phrases (almost half of this kind of phrase) as well as most of the pleading phrases.

Cluster 2 exhibits the lowest ratios across all analysed parameters. It is the intonationally least expressive cluster but the most frequent, comprising 65% of the phrases.

Cluster 3 is characterised by the highest intensity, while the other parameters differ little from those of Cluster 1. Only 7% of phrases belong to this cluster; these consist of isolated command and pleading phrases and one phrase with the performative verb *siūlau* 'I suggest'.

Table 3. Parameters of cluster centres

Feature	Cluster		
	1	**2**	**3**
Mean F0	1,03	0,98	1,02
F0 maximum	1,33	1,31	1,34
F0 range	1,03	0,99	1,04
F0 range centre	1,01	1,00	0,98
Mean intensity	1,98	0,72	3,52
Intensity maximum	10,02	3,45	20,20
Duration of *perrašyk darbą / perrašyti darbą*	1,07	1,00	1,04
Number of phrases (%)	27	65	8

Table 4. Distribution of phrases across clusters

Phrase	Cluster			Phrase	Cluster		
	1	**2**	**3**		**1**	**2**	**3**
IP1		20		IP9	4	16	
IP2	10	9	1	IP10	2	18	
IP3	1	19		IP11	16	1	3
IP4		20		IP12	11	7	2
IP5	4	12	4	IP13	7	5	8
IP6	5	15		IP14		19	1
IP7	7	13		IP15	3	17	
IP8	12	4	4				

Almost three quarters (72%) of request phrases belong to the least expressive Cluster 2, 25% are assigned to Cluster 1, and only individual cases fall into Cluster 3. The vast majority (86%) of advice phrases are assigned to Cluster 2 and 13% to Cluster 1. Command phrases are widely distributed: 44% in Cluster 1, 37% in Cluster 2, and 19% in Cluster 3. These results indicate that the degree of acoustic expressiveness in phrases expressing active directives varies; they diverge more or less from phrases that have relatively neutral characteristics.

4. Conclusions

The study can be summarised as follows (95% confidence intervals for the median ratios are given in brackets):

1. Mean F0: IPreq (0.97–1.01) – IPcomm (0.96–1.01) – IPadv (0.97–1.03),
2. F0 maximum: IPreq (1.25–1.31) – IPcomm (1.24–1.33) – IPadv (1.32–1.42),
3. F0 range: IPreq (0.90–1.00) – IPcomm (0.90–1.04) – IPadv (1.01–1.17),
4. F0 range centre: IPreq (0.96–1.00) – IPcomm (0.96–1.01) – IPadv (1.00–1.05),
5. Mean intensity: IPreq (1.00–1.37) – IPcomm (1.72–2.18) – IPadv (0.60–0.76),
6. Intensity maximum: IPreq (4.54–6.10) – IPcomm (8.77–11.18) – IPadv (3.47–4.69),
7. Duration of *perrašyk darbą / perrašyti darbą*: IPreq (1.07–1.13) – IPcomm (0.99–1.03) – IPadv (0.94–0.97).

The F0 parameters are not straightforward: mean F0 does not show any clear tendencies as the intervals for all phrases overlap. Higher F0 maximum, a wider F0 range, and a higher range centre are characteristic of advice. The intonation contour is either falling or rising–falling in examples with two intermediate phrases.

Intensity appears to be a reliable indicator of active directives, such as commands. Advice phrases consistently show the lowest intensity. It must be acknowledged, however, that the force of a directive is not easily and unambiguously defined. While it is natural to describe commands as stronger than requests or advice when comparing directive subtypes, the results of this study demonstrate that, for instance, a pleading phrase may exhibit more prominent acoustic properties than certain commands.

A longer than usual duration signals a request. In the material, both syllable nuclei and consonants are lengthened in request phrases. Advice phrases are the shortest in the

present data. However, based on this study alone it cannot be concluded that they are generally shortened, as a more detailed analysis of other functional phrase types – particularly statements – is required.

The results of the cluster analysis indicate that, based on a combination of acoustic parameters, it is possible to reliably distinguish between active and less active directives. However, the specific nature of the latter – advice, suggestion, desire, request, invitation to act, etc. – remains difficult to identify.

The comparison of phrases with the same phonemic structure leads to the following conclusions:

1. Expressively produced phrases differ from the other examples in that all F0 parameters are elevated, resulting in a generally high intonation contour, and intensity is high. They are longer than phrases of the same phonemic structure produced as commands or advice, but their duration differs little from that of request phrases. These phrases additionally exhibit a large dispersion in intensity and duration values, indicating that these parameters may deviate considerably from the habitual speaking characteristics.
2. The results for non-expressively produced phrases vary, but certain regularities can be observed:
   2.1. Command phrases are characterised by quite high intensity. Their F0 maximum and range are the smallest, whereas the mean F0 occupies an intermediate position between request and advice. They are shorter than request phrases but differ little in duration from advice phrases. Compared with other non-expressive phrases, commands are distinguished by a smoother intonation contour and a somewhat higher final F0, however, they do not exhibit a rising intonation contour.
   2.2. Request phrases differ from other non-expressive phrases primarily in the longer duration. Their mean F0 is the highest of all three types, whereas F0 maximum, ranges, and intensity occupy an intermediate position between commands and requests. They begin with a high F0, which accounts for the elevated mean F0.
   2.3. Advice phrases show low intensity. Their mean F0 is the lowest, but maximum and ranges are relatively high. They are shorter than request phrases but differ little in duration from commands. It may be hypothesised that these phrases are closest to the speaker's habitual speaking properties.

The comparison of phrases with performative verbs in the subjunctive and present tense yields ambiguous results:

a) Commands and advice tend to be produced with higher F0 when expressed with performatives in the subjunctive, whereas requests, conversely, tend to exhibit higher F0 in the present tense.
b) Requests and advice expressed in the subjunctive are produced with higher intensity, whereas commands are more intense when using verbs in the present tense.

Thus, the study partially supports E. Gudavičienė's (2006, 65) conclusion that the subjunctive denotes a possible or desired action and is therefore employed to express a more gentle and polite request or advice.

The study does not provide a clear answer as to whether acoustic properties are merely complementary when directives are expressed with performative verbs. If the acoustic features were more prominent in phrases with performatives, it could be

predicted that the verb emphasises the nature of the directive. Conversely, if the acoustic features were weaker, it would suggest that the directive is conveyed lexically rather than through acoustic means.

The F0 of request phrases is higher in phrases without a performative verb, except in cases of pleading. Commands and advice, however, tend to be produced with higher F0 when a performative verb is present. Requests and advice were produced more intensely when a performative verb was included, whereas the intensity of commands did not show a clear pattern and remained similar across different structures. Since, according to the results of this study, intensity is a reliable indicator of directive force, it can be tentatively assumed that the verb strengthens requests and advice, while commands are maintained at the same intensity level. Considering that phrases without performative verbs were not well recognised, it may cautiously be concluded that the verb not only reveals the type of directive and strengthens directive force but also highlights the acoustic properties.

Acknowledgment

This research was funded by a grant (No. S-MIP-24-90) from the Research Council of Lithuania (Project "Intonation of the Lithuanian language: Functional Types and Interference with Prosodic Units"). Sincere thanks are due to the actors whose recordings were used in the analysis, as well as to the colleague who assisted with the recordings in the studio. Special thanks go to the phoneticians who participated in the survey.

References

Balkevičius, J. (1963). *Dabartinės lietuvių kalbos sintaksė* (Syntax of the modern Lithuanian language), Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla.

Bikulčienė, P. (1978). Skatinimo ir konstatavimo intonacijų gretinimas (Comparison of intonation of directives and statements), *Kalbos garsai ir melodika*, 3–11.

Boersma, P., Weenink, D. (2018). *Praat: doing phonetics by computer, version 6.035*. www.praat.org.

Cruttenden, A. (1997). *Intonation*, 2nd ed., Cambridge: Cambridge University Press.

Drukteinis, A. (2024). Sintaksė (Syntax), in A. Balčiūnienė, A. Drukteinis, R. Kazlauskaitė, J. Vaskelienė, *Praktinė bendrinės lietuvių kalbos gramatika* (Practical grammar of the Lithuanian language), Vilnius: Vilniaus universiteto leidykla, 489–755.

Falé, I., Faria, I. H. (2007). Imperatives, orders and requests in European Portuguese intonation. Proceedings of the *16th International Congress of Phonetic Science*s, 1041–1044.

Féry, C. (2017). *Intonation and Prosodic Structure*, Cambridge University Press, Cambridge.

Frota, S., Prieto, P. (2015). Intonation in Romance: systemic similarities and differences, in S. Frota, P. Prieto (eds.), *Intonation in Romance*, Oxford: Oxford University Press, 392–418.

Gårding, E. (1998). Intonation in Swedish, in D. Hirst, A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge: Cambridge University Press, 112–130.

Gibbon, D. (1998). Intonation in German, in Daniel Hirst, Albert Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge: Cambridge University Press, 78–95.

Gudavičienė E. (2006). Direktyvai kaip ilokucinių aktų rūšis (Directives as a kind type of illocutionary acts). *Lituanistica* **67**(3), 60–68.

IBM Corp. (2020). *IBM SPSS statistics for Windows*, version 27.0, computer software, IBM Corp.

Kazlauskienė, A., Dereškevičiūtė, S. (2022). Observations on the prosodic marking of narrow focus in Lithuanian. *Baltic J. Modern Computing*, **10**(3), 307–316. doi:10.22364/bjmc.2022.10.3.04

Kazlauskienė, A., Dereškevičiūtė, S., Sabonytė, R. (2023). *Bendrinės lietuvių kalbos intonacija: frazės centras, ribos ir žymėjimas* (Standard Lithuanian intonation: phrase centre, boundaries, and marking), Kaunas: Vytauto Didžiojo universiteto leidykla, available at https://doi.org/10.7220/9786094675782.

Kundrotas, G. (2018). *Lietuvių kalbos intonacinė sistema* (The intonation system of the Lithuanian language), Utena: Indra.

Labutis, V. (1998). *Lietuvių kalbos sintaksė* (Lithuanian syntax), Vilniaus universiteto leidykla, Vilnius.

Mikalauskaitė, E. (1975). *Lietuvių kalbos fonetikos darbai* (Works on Lithuanian phonetics), Vilnius: Mokslas.

Oginskienė, E. (1994). Modaliniai sakinių tipai (Modal sentence types), Ambrazas, V. (ed.), *Dabartinės lietuvių kalbos gramatika* (Grammar of the modern Lithuanian language), Mokslo ir enciklopedijų leidybos institutas, Vilnius, 574–581.

Rietveld, T., Vermillion P. (2004). Cues for Perceived Pitch Register. *Phonetica: international journal of phonetic science*. **60**(4), 261-272.

Sirtautas, V., Grenda, Č. (1988). *Lietuvių kalbos sintaksė* (Lithuanian syntax), Vilnius: Mokslas.

Vaitkevičiūtė, V. (1964). Kai kurios skatinimo rūšys ir jų reiškimo būdai lietuvių kalboje (Some types of encouragement and their expressions in the Lithuanian language), *Lietuvos TSR Mokslų Akademijos darbai*, serija A, **1**(16), 219–229.

Vaitkevičiūtė, V. (1966). Lietuvių literatūrinės kalbos klausiamieji sakiniai (Interrogative sentences in Lithuanian Standardt language), *Lietuvos TSR Mokslų Akademijos darbai*, serija **A**, 1(20), 193–202.

Xu, Y. (2013). ProsodyPro – A tool for large-scale systematic prosody analysis, in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, 7–10.

Received January 13, 2026, accepted January 25, 2026

Baltic J. Modern Computing, Vol. 14 (2025), No. 1, 50-69
https://doi.org/10.22364/bjmc.2026.14.1.03

ADF^{2}S Feature Set of Anomaly Detection Framework Based on Knowledge Discovery

Imelda ZADEJA

Vidzeme University of Applied Sciences, Terbatas Street 10, Valmiera, Latvia

imelda.zadeja@va.lv

ORCID 0009-0000-7305-8332

Abstract. In the modern knowledge era, the exponential growth of digital solutions has led to the generation of vast amounts of data. This necessitates the development of data and knowledge-driven advanced techniques to extract insights and support informed decision-making. Within this context, the detection of anomalies, data points that deviate significantly from expected patterns, becomes crucial as these anomalies can arise due to diverse factors, including sensor errors, data corruption, and changes in underlying processes, all of which may impact system performance, accuracy, and overall efficiency. This paper thoroughly examines the diverse frameworks and architectures established for anomaly detection across multiple domains. It highlights the complexity associated with the nature of anomalies, which are often domain-specific and contextually bound, thereby presenting significant challenges in devising a universal framework capable of addressing anomalies regardless of the domain or context of the application. To address these challenges, the author proposes a comprehensive anomaly detection framework feature set (ADF^{2}S) that captures the functional, structural, and operational dimensions of anomaly detection frameworks. The proposed anomaly detection framework feature set (ADF^{2}S) and its cross-framework evaluation contribute a practical foundation for researchers and practitioners, supporting the development of anomaly detection frameworks capable of balancing scalability, interpretability, and resilience.

Keywords. anomaly detection, anomaly detection framework, knowledge discovery, machine learning, anomaly detection framework feature set (ADF^{2}S).

1. Introduction

In an era of exponential growth of data, the ability to detect anomalous behaviour within large and complex datasets has become more critical than ever. From cybersecurity threats and fraud detection to industrial system failures and health monitoring, anomaly detection serves as a foundational technique for ensuring reliability, security, and efficiency across domains. Anomalies, also known as outliers, novelties, or exceptions, represent patterns in data that deviate significantly from expected behaviour (Chandola et al., 2009). Their identification is often an early indication of critical events or hidden knowledge (Pimentel et al., 2014). Anomaly detection is naturally related to the broader field of knowledge discovery. As a central step in the knowledge discovery process,

anomaly detection enables the extraction of actionable and previously unknown insights from massive datasets. In this context, machine learning plays a crucial role by providing scalable, adaptable, and intelligent models capable of identifying complex and subtle deviations from norms (Aggarwal, 2013). Whether through supervised, unsupervised, or semi-supervised approaches, machine learning techniques are frequently at the core of modern anomaly detection frameworks. The importance of anomaly detection has grown rapidly with the increasing volume, velocity, and variety of data generated in real-time from various systems.

The presence of sensors, interconnected systems, and user-driven digital platforms has led to environments where millions of data points are produced every second. This data presents unique challenges, including the high dimensionality of data, concept drift, class imbalance, and lack of labelled data, particularly for rare or unknown anomalies (Goldstein and Uchida, 2016). These difficulties necessitate sophisticated frameworks capable of adaptive learning, contextual awareness, and explainability. Beyond technical complexity, anomaly detection also faces conceptual and operational challenges. The very definition of what constitutes an "anomaly" is often domain-specific and context-dependent (Chandola, et al., 2009). In some scenarios, anomalies may reflect genuine faults; in others, may reveal novelties or trends (Pimentel et al., 2014). Thus, frameworks must be precise in identifying irregularities and flexible enough to interpret their significance in a dynamic environment (Ahmed et al., 2015). New frontiers are being explored in areas such as explainable AI and graph-based anomaly detection, which integrate expert feedback into the learning process (Choi et al.,2022) (Eberle and Holder, 2015). As such, anomaly detection is an actual research area that adapts to emerging technological, societal, and methodological shifts.

2. From knowledge to knowledge discovery: foundations for anomaly detection

In the context of modern data-driven systems, the concept of knowledge extends far beyond raw information. Knowledge can be broadly defined as structured, contextualised, and actionable information, which enables informed decision-making, problem-solving, and innovation (Nonaka and Takeuchi, 1995). Unlike data, which consist of unprocessed facts, or information, which refers to organized data imbued with meaning, knowledge reflects an integration of experience, interpretation, and insight. With the digital transformation of industries and services, knowledge workers - individuals who engage in tasks involving critical thinking, analysis, and problem-solving - have become central to organizational value creation (Davenport, 2005). As data volume and complexity increase, manual or intuitive methods for deriving knowledge become insufficient, paving the way for computational support in uncovering patterns and relationships (Jansevskis and Osis, 2024). The paradox of today's digital ecosystem lies in the fact that while organisations are inundated with data, they often lack knowledge. Extracting useful, actionable insights from terabytes or petabytes of data demands systems that go beyond storage and computation, require systems that understand the semantics of information and can detect deviations from expected patterns (Thudumu et al., 2020). This is where knowledge discovery becomes indispensable.

Knowledge Discovery in Databases (KDD) refers to the process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et al., 1996). This process focuses on transforming data into knowledge through a sequence of operations, including data selection, preprocessing, transformation, data mining, and interpretation/evaluation (Frawley et al., 1992). Within this process, data mining plays a crucial role by applying algorithms to extract patterns from data (Al-Faouri, 2023). In this context, anomaly detection is one of the core data mining tasks, alongside classification, clustering, and association rule learning (Klosgen and Zytkow, 1996). In contrast to traditional analytical approaches, knowledge discovery enables systems to identify unexpected behaviours, emergent trends, and hidden structures, which are often indicated by anomalies.

3. Overview of anomaly detection

Anomaly detection has been an intriguing field for researchers and practitioners that has been studied for centuries (Nassif et al., 2021). Numerous distinct methods and approaches have been developed over time to detect anomalies across different applications, domains, and contexts. Starting from statistical methods to machine learning and neural networks, anomaly detection methods have been developed and reshaped by employing sophisticated and complex methods that improve the performance and efficiency of such systems (Mishra and Kumar, 2022). The concept of anomalies is defined as patterns in data that do not conform to a well-defined notion of expected behaviour (Barnard and Stryker, 2023). As such, patterns or behaviours in data that deviate from normal behaviour are referred to as anomalies or outliers, depending on the domain or context in which anomaly detection techniques are applied (Steenwinckel et al., 2021) (Injadat et al., 2018). Examples of such applications include fraud detection, loan application processing, monitoring of medical conditions, cybersecurity intrusion detection, and fault detection for predictive maintenance (Nassif et al., 2021).

4. Anomaly detection frameworks

An essential consideration in the development of anomaly detection frameworks is the acknowledgement of the domain-specific nature of anomalies (Alam et al., 2019). This characteristic introduces significant challenges in creating a generic framework capable of detecting all forms of anomalies, regardless of domain-specific conditions or data characteristics (Pang et al., 2021). The complexity associated with high-dimensional data and contextual anomalies cannot be understated, as they prevent the generalizability of detection approaches across different applications (Choi et al., 2021). Moreover, there is a need for a comprehensive taxonomy that systematically analyses the diverse anomaly detection frameworks' features and capabilities based on their characteristics and level of granularity (Feng et al., 2025) (Zhou et al., 2022).

Anomaly detection can signal system faults, cyberattacks, or inefficiencies, making their timely and accurate detection crucial for operational resilience and efficiency (Cauteruccio et al., 2021) (Fan and You, 2024). Over the past decade, anomaly detection frameworks have evolved from rule-based approaches to sophisticated machine learning

driven systems capable of processing complex, high-dimensional, and temporal data (Chai et al., 2022). Modern anomaly detection frameworks and architectures often integrate components, including feature selection, model ensemble, time-series analysis, and explainability. These components are crucial for identifying relevant patterns and attributing anomalous behaviour to specific systemic or operational contexts. The effectiveness of such frameworks largely depends on how they leverage selected features, whether statistical, contextual, or domain-specific, to inform the detection process (Alam et al., 2019).

In this research, three influential anomaly detection frameworks are analysed in depth to examine how each leverages selected features for anomaly identification. These frameworks were chosen following a systematic review of 54 anomaly detection frameworks identified through searches of scientific databases (Scopus, ScienceDirect, and IEEE Xplore) using the terms "anomaly detection framework" and "anomaly detection architecture". Framework selection was guided by the following criteria: [1] the framework was published within the past ten years; and [2] the corresponding publication achieved a Field-Weighted Citation Impact (FWCI) of at least five according to Scopus, indicating a substantial contribution to the domain. In addition to these criteria, the selected frameworks represent distinct architectural paradigms and methodological diversity, thereby providing a comprehensive cross-section of the anomaly detection landscape. Based on these criteria, the following state-of-the-art anomaly detection frameworks were selected: "Opprentice: Towards Practical and Automatic Anomaly Detection Through Machine Learning" (Liu et al., 2015), "An Ensemble Learning Framework for Anomaly Detection in Building Energy Consumption" (Araya et al., 2017), and "Time-Series Anomaly Detection Service at Microsoft" (Ren et al., 2019). Collectively, these frameworks serve as representative examples of the broader design space, encompassing supervised and unsupervised learning, ensemble-based modelling, time-series anomaly detection, domain-specific integration, and production-grade scalability. Moreover, they have demonstrated significant impact in both academia and industry, as reflected in their citation volume, adoption, and influence on subsequent research and industrial solutions.

4.1. Opprentice: automatic anomaly detection through machine learning

Among the foundational contributions to the field of anomaly detection, the framework depicted in Figure 1, proposed in the research article "Opprentice: Towards Practical and Automatic Anomaly Detection Through Machine Learning", stands out for its practical orientation and system-level applicability (Liu et al., 2015). Unlike traditional methods that often require expert intervention for algorithm selection and parameter tuning, Opprentice introduces an end-to-end, automated anomaly detection system that leverages a meta-learning strategy to dynamically select the most appropriate detection algorithm based on the characteristics of the input data (Liu et al., 2015). This feature enhances usability and lowers the technical barrier for deployment in real-world monitoring systems. The framework is structured around several key components: a data collector and repository, a feature extractor, a model training and selection engine, and a runtime detector. Together, these components enable the system to process large volumes of heterogeneous system metrics, train and select the most suitable detection model, and

then apply this model for real-time anomaly detection with minimal computational overhead.

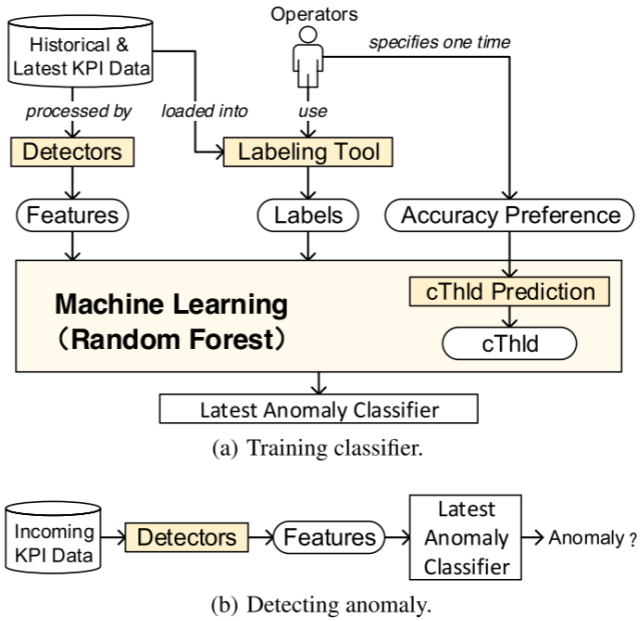


Figure 1: Opprentice anomaly detection architecture (Liu et al., 2015)

The framework is particularly relevant in large-scale IT operations and cloud environments, where telemetry data is generated continuously and must be processed in real time (Liu et al., 2015). By validating the approach on extensive production datasets, the study demonstrates high detection accuracy and low false-positive rates, thereby offering a scalable and efficient solution for anomaly detection in operational settings. The Opprentice anomaly detection framework contributes significant value to the domain by bridging the gap between theoretical models and practical, large-scale deployments. Moreover, it has influenced the design of subsequent industrial-grade detection systems, underscoring its enduring relevance and foundational role in the evolution of automated, machine learning–based anomaly detection frameworks.

4.2. Ensemble anomaly detection framework in energy consumption

The second selected framework in Figure 2, developed focuses on smart buildings, where energy consumption data presents a challenging time series prediction problem. In their work, the authors proposed an Ensemble Anomaly Detection (EAD) framework specifically designed to detect anomalous patterns in building energy consumption data (Araya et al., 2017). The EAD framework is based upon the core principles of Collective Contextual Anomaly Detection using the Sliding Window (CCAD-SW) framework,

adapted by the authors. The results indicated that the EAD framework improved the sensitivity of the CCAD-SW by 3.6% and reduced the false alarm rate by 2.7% (Araya et al., 2017). As a result, it demonstrates the effectiveness of the ensemble approach in enhancing anomaly detection capabilities.

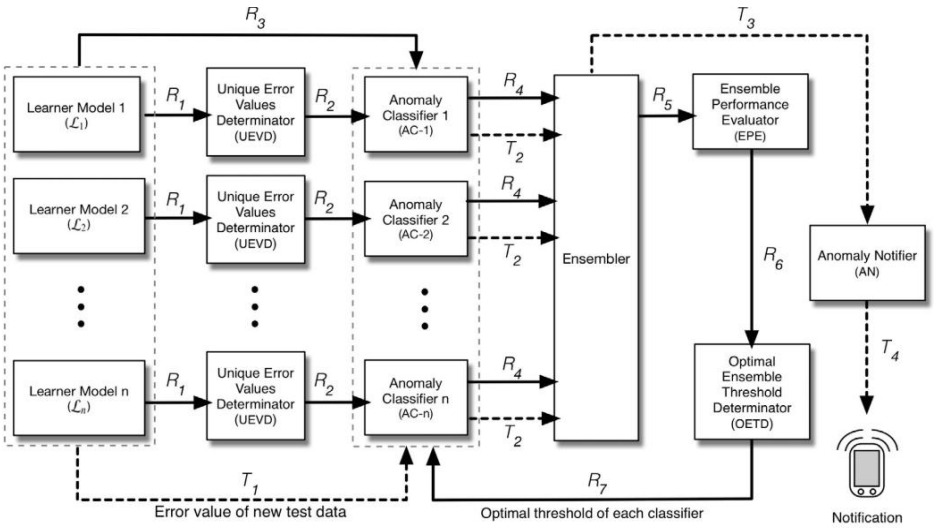


Figure 2: Ensemble anomaly detection (EAD) framework (Araya et al., 2017)

The primary aim of the framework is to enhance the accuracy and robustness of anomaly detection in the presence of complex, multivariate, and temporally dynamic energy usage profiles. The EAD framework employs a heterogeneous ensemble learning approach that integrates multiple base learners, including One-Class Support Vector Machines (OC-SVM), Isolation Forests, and Autoencoders, each capturing distinct characteristics of the data (Araya et al., 2017). By aggregating the outputs of these individual models, the ensemble produces a composite anomaly score that is more resilient to the limitations of any single method (Araya et al., 2017). These predictive components are trained to estimate expected energy consumption based on historical data and contextual variables. Anomalies are then inferred from deviations between predicted and observed values. The EAD framework enhances robustness and reliability by combining the outputs of pattern-based and prediction-based models using an ensemble approach. It is found that the optimal detection threshold for the ensemble is the result of an independent optimization process, emphasizing the importance of ensemble calibration.

4.3. Time series anomaly detection service at Microsoft

In this research, the authors present a comprehensive framework for time-series anomaly detection that represents a production-scale system deployed at Microsoft, as illustrated in Figure 3. The framework is designed for real-time monitoring of time-series data across a range of cloud-based applications, including Office and Azure, and emphasizes

scalability, automation, and low latency (Ren et al., 2019). To achieve these objectives, the authors propose a novel algorithm that combines Spectral Residual (SR) and Convolutional Neural Network (CNN) techniques. This work represents the first application of the SR model, originally developed for visual saliency detection, to the domain of time-series anomaly detection (Ren et al., 2019). The integration of SR and CNN significantly enhances detection performance, leading to substantial improvements in anomaly detection accuracy and robustness.

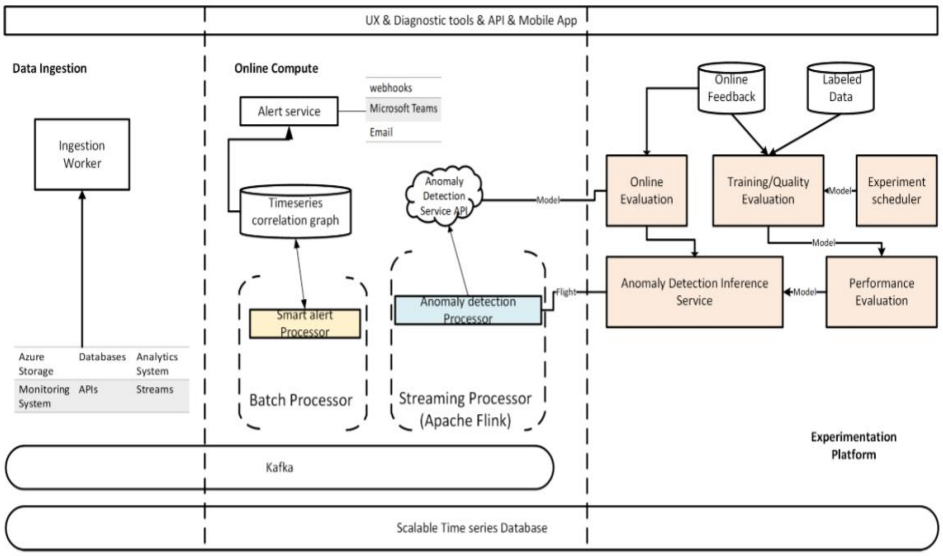


Figure 3: Time series anomaly detection system at Microsoft (Ren et al., 2019)

The proposed framework outperforms existing state-of-the-art models, achieving an F1-score improvement of over 20% on Microsoft production data (Ren et al., 2019), demonstrating the effectiveness of the combined approach in real-world applications. It is designed to be general and efficient, enabling seamless integration with online monitoring systems and providing rapid alerts for critical metrics, essential for large-scale, real-time operations. The framework maintains an unsupervised learning approach by generating synthetic anomalies to train the CNN. In practical deployments, it has allowed product teams to detect issues more quickly, reduce manual intervention, and accelerate diagnostic processes.

5. Anomaly detection framework feature set

The design and evaluation of anomaly detection frameworks require a structured understanding of the core capabilities these systems must possess. Based on a comprehensive review of the literature and the analysis of existing state-of-the-art frameworks mentioned in the previous section, the author proposes a feature set that

captures the most crucial dimensions along which anomaly detection frameworks can be characterised. As shown in Figure 4, the proposed Anomaly Detection Framework Feature Set (ADF²S) includes: Data Handling & Input Capabilities, Anomaly Detection Type Support, Detection Techniques Flexibility, Adaptability, Explainability and Interpretability, and Data Quality Aware Anomaly Detection. Each feature of the ADF²S is further divided into subcategories, which represent specific mechanisms, data properties, or learning paradigms relevant to the main feature.

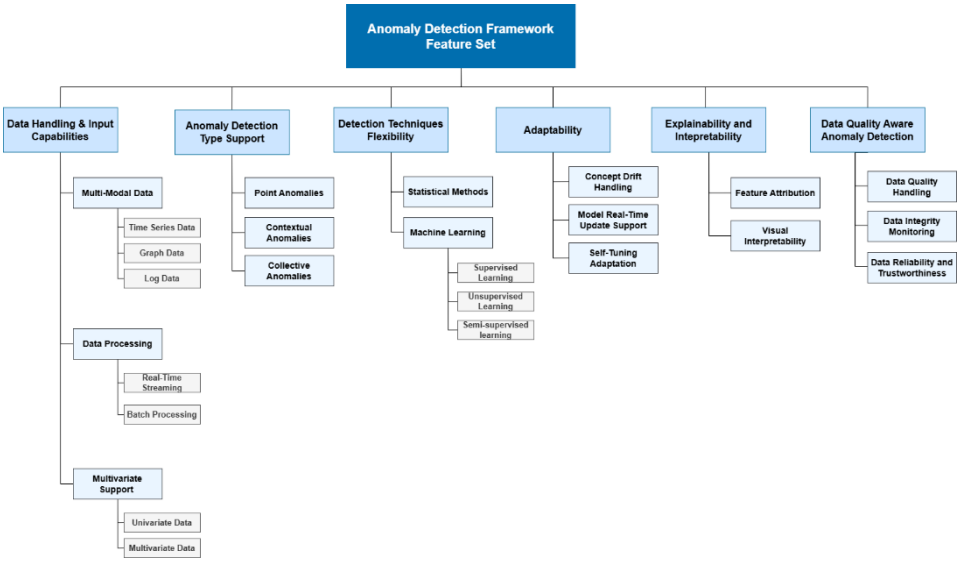


Figure 4: Anomaly detection framework feature set (ADF²S)

The primary categories in ADF²S capture high-level functional dimensions, such as data handling, detection techniques, and adaptability, because they reflect broad system-level requirements that recur across domains. However, these high-level capabilities are too coarse to meaningfully differentiate between frameworks. To address this, each main feature is decomposed into sub-features that represent the concrete technical mechanisms, algorithmic paradigms, or system design choices through which the high-level capability is operationalized. The hierarchical structure aligns with the way this research organizes anomaly detection methods and system capabilities. Grouping conceptually related mechanisms under a shared parent feature ensures that the taxonomy maintains conceptual coherence and follows established scientific conventions. Overall, the feature set taxonomy provides a conceptual platform for evaluating the completeness, flexibility, and domain suitability of anomaly detection frameworks, enabling a nuanced discussion of how design choices affect system performance and applicability.

The review followed an iterative synthesis approach: first, a broad corpus of anomaly detection studies, including machine learning–based approaches, time-series models, and system-level architectures, was examined to identify recurring functional requirements and modelling practices. These concepts were then cross-referenced with capabilities described in research articles and industry-grade platforms to ensure practical relevance and contemporary applicability. Features were grouped and refined through thematic clustering, allowing coherent high-level categories and their corresponding sub-features to emerge. This methodology ensured that the resulting ADF^{2}S taxonomy is grounded in scientific knowledge while reflecting operational considerations observed in real-world anomaly detection deployments. Consequently, the feature set represents an evidence-based synthesis of the key conceptual and technical elements consistently highlighted across both academic literature and industrial implementations.

5.1. Data handling and input capabilities anomaly detection feature

The capability of an anomaly detection framework largely depends on its ability to effectively handle and process the data on which it operates. The data handling and input capability feature captures a framework's flexibility in accommodating diverse data types, processing paradigms, and levels of structural complexity (Goldstein and Uchida, 2016). This feature is further subdivided into three components: multi-modal data support, data processing, and multivariate support, each of which plays a critical role in defining the framework's operational scope and applicability across domains.

Modern systems generate data from heterogeneous sources, including sensors, logs, network traffic, and structured time-series metrics. Anomaly detection frameworks that support multi-modal data can integrate insights across different data representations, enabling a more holistic understanding of system behaviour. In this taxonomy, multi-modal data is categorized into three subtypes: time-series data, graph data, and log data. Time-series data is the most prevalent input format for anomaly detection, particularly in domains such as industrial monitoring, finance, and healthcare (Choi et al., 2021) (Thudumu et al., 2020). Frameworks must process temporally ordered observations to capture trends, seasonality, and temporal anomalies (Ren et al., 2019).

Graph-structured data, increasingly important in cybersecurity, social networks, and network infrastructure, encodes relationships among entities. Frameworks supporting graph data can identify structural or topological anomalies, such as compromised nodes or anomalous information flows (Eberle and Holder, 2015). Log data, common in IT and software systems, consists of unstructured or semi-structured textual records. Effective anomaly detection frameworks parse, tokenize, and extract patterns from logs to detect deviations in event sequences or semantic content (Filzmoser et al., 2008).

Anomaly detection frameworks also differ in their data ingestion and analysis strategies, which are broadly categorized as batch processing and real-time streaming. Batch processing involves analysing data in fixed intervals or accumulated batches and is well-suited for retrospective analysis, scheduled system checks, or environments where immediate response is not critical (Xu et al., 2023). Due to its computational simplicity and ease of implementation, this approach is commonly adopted by traditional frameworks (Araya et al., 2017). In contrast, real-time streaming enables continuous processing of incoming data streams, allowing anomalies to be detected as they occur. Such capabilities are essential in domains requiring near real-time responses, including

financial fraud detection, cloud service monitoring, and fault detection in critical infrastructure (Ren et al., 2019).

In complex systems, anomalies often arise from interactions among multiple variables rather than from isolated data streams. The multivariate support sub-feature captures a framework's ability to model these interdependencies (Wang et al., 2025). Univariate approaches analyse each data stream independently and, while computationally efficient, may fail to detect correlated anomalies spanning multiple variables. Multivariate approaches, by contrast, analyse multiple data streams simultaneously, enabling the detection of more complex and interrelated anomaly patterns. Collectively, these data handling and input capabilities define how effectively an anomaly detection framework can operate across varying data types, processing requirements, and system complexities.

5.2. Anomaly detection type support anomaly detection feature

The anomaly detection type support feature characterizes the types of anomalies a framework is capable of identifying, which are closely related to the nature of the underlying data. Broadly, anomalies are commonly classified into three categories: point anomalies, contextual anomalies, and collective anomalies (Chandola et al., 2009). A point anomaly occurs when an individual data instance significantly deviates from the rest of the dataset and is therefore considered anomalous. This is the simplest and most commonly addressed form of anomaly. Examples include a sudden spike in a patient's blood pressure or an unusually large transaction amount in a banking system. A contextual anomaly arises when a data instance is anomalous only within a specific context, despite appearing normal in other circumstances (Chandola et al., 2009). In such cases, contextual information, such as time, location, or environmental conditions, is essential for correct interpretation. For example, a temperature of 10 °C may be normal during winter but anomalous during summer in a given region. Collective anomalies refer to groups of related data instances that are anomalous when considered together, even though individual instances within the group may not appear anomalous on their own (Araya et al., 2017). These anomalies often emerge from sequential or correlated patterns in the data. A typical example is a series of login attempts from different IP addresses within a short time window, which may indicate a coordinated attack (Choi et al., 2021). Because collective anomalies are not easily identifiable at the level of individual data points, long-term contextual and temporal analysis is particularly important for their detection. Consequently, support for different anomaly types constitutes a critical characteristic of anomaly detection frameworks. It influences the choice of detection techniques, contextual modelling strategies, and overall system design, thereby aligning framework capabilities with the specific anomaly patterns present in the target application domain.

5.3. Detection techniques flexibility anomaly detection feature

The detection technique flexibility feature captures the range of methods and algorithms employed by anomaly detection frameworks across different system architectures. Statistical anomaly detection techniques represent some of the earliest approaches in this domain (Chandola et al., 2012). These methods model normal data behaviour using

statistical assumptions and identify anomalies as deviations from the learned distribution. Common statistical approaches include parametric techniques such as Gaussian distribution modelling, hypothesis testing, and time-series models (Bhuyan et al., 2014). In recent decades, machine learning–based techniques have become increasingly prominent in anomaly detection. Machine learning aims to automate knowledge acquisition from data examples (Bose and Mahapatra, 2001) and is widely used to construct models that distinguish between normal and anomalous patterns (Mahapatra and Bose, 2001). Advanced machine learning algorithms are particularly effective in handling high-dimensional data and learning complex patterns, thereby improving detection accuracy and efficiency (Herr et al., 2021).

Based on the availability of labelled data, anomaly detection techniques are commonly categorized into supervised, unsupervised, and semi-supervised approaches. Supervised anomaly detection relies on labelled datasets containing both normal and anomalous instances. In this setting, a predictive model is trained to discriminate between the two classes. However, supervised approaches face several challenges, including the severe imbalance between normal and anomalous samples and the difficulty of obtaining accurate and representative labels for anomalies (Chatterjee and Ahmed, 2022). Additionally, discrepancies between training and testing anomaly distributions can degrade performance. In contrast, unsupervised anomaly detection does not require labelled training data and operates under the assumption that anomalies are rare compared to normal instances. These methods identify deviations based on inferred data structure or distribution. While widely applicable, unsupervised approaches may produce high false alarm rates when the underlying assumptions about anomaly rarity or data distribution do not hold (Nassif et al., 2021). Consequently, much recent research has focused on improving unsupervised techniques, particularly for large-scale and unlabelled datasets (Pang et al., 2021).

Semi-supervised anomaly detection occupies a middle ground, where training data consists only of labelled normal instances. Any data point that deviates from the learned normal behaviour is flagged as anomalous. Because anomaly labels are not required, semi-supervised methods are often more practical than fully supervised approaches. Many semi-supervised techniques can also be adapted to operate in an unsupervised manner by assuming that training data contains very few anomalies, which minimally affect model learning (Nassif et al., 2021). Across application domains such as finance, manufacturing, and the Internet of Things (IoT), supervised, unsupervised, and semi-supervised machine learning approaches have been extensively studied and applied (Torr and Murray, 1993) (Marsland, 2001). Recent studies suggest that semi-supervised methods often outperform supervised approaches in real-world settings, particularly when labelled anomaly data are scarce or incomplete (Sunny et al., 2022). These techniques are especially suitable for high-dimensional datasets with large proportions of unlabelled data, a common characteristic in many operational environments (Filzmoser et al., 2008).

5.4. Adaptability anomaly detection feature

The adaptability feature captures a framework’s ability to maintain accuracy and relevance in dynamic environments where data distributions evolve. In real-world systems, such as cloud infrastructure and financial markets, the assumption of

stationarity rarely holds (Chatterjee and Ahmed, 2022). Consequently, anomaly detection models must incorporate mechanisms to respond to shifting behavioural patterns, commonly referred to as concept drift, to prevent performance degradation over time (Chatterjee and Ahmed, 2022). This dimension comprises three interrelated sub-features that collectively define a framework's adaptability: concept drift handling, model real-time update support, and self-tuning adaptation.

Concept drift refers to gradual changes in the underlying data distribution that alter the definition of normal behaviour. If left unaddressed, such changes can lead to increased false positives or missed anomalies. Frameworks that support concept drift handling can detect these shifts and adjust their internal models accordingly, for example, by recalibrating decision boundaries, retraining models, or updating statistical parameters (Chatterjee and Ahmed, 2022).

Model real-time update support denotes a framework's ability to incrementally update its models in response to incoming data, rather than relying on full retraining. This capability is critical in real-time applications where batch retraining is computationally expensive or insufficiently responsive to rapid system changes (Zuo et al., 2019). In anomaly detection, real-time updates enable continuous model refinement as new patterns emerge, allowing for rapid adaptation. Frameworks with this capability are particularly well suited to streaming data environments, such as network monitoring and predictive maintenance in sensor networks (Choi et al., 2021).

Self-tuning adaptation refers to internal mechanisms that automatically adjust detection parameters, such as thresholds, sensitivity levels, or window sizes, without manual intervention (Alam et al., 2019). This capability reduces reliance on domain expert tuning, thereby improving scalability and deployment efficiency. By adapting to variations in noise levels, system load, and behavioural patterns, self-tuning mechanisms help anomaly detection frameworks maintain robust performance across changing operational conditions.

5.5. Explainability and interpretability anomaly detection feature

In modern anomaly detection frameworks, explainability and interpretability have emerged as essential features, particularly in safety-critical and regulated application domains (Zhong et al., 2023). These capabilities describe the extent to which a model's decision-making process, specifically its anomaly detection outcomes, can be understood, traced, and justified by human users. As anomaly detection systems are increasingly integrated into operational pipelines, the ability to explain why a particular data point is classified as anomalous is critical for establishing trust, ensuring accountability, and enabling informed responses (Choi et al., 2022). This feature dimension comprises two primary subcomponents: feature attribution and visual interpretability, each contributing in distinct ways to user understanding of the detection process.

Feature attribution refers to a framework's ability to identify and quantify the influence of individual input features on an anomaly detection decision. In machine learning–based systems, particularly those employing ensemble or deep learning methods, detection mechanisms are often opaque (Zhong et al., 2023). Feature attribution techniques mitigate this opacity by producing saliency scores or importance rankings that highlight which variables or input patterns contributed most strongly to an

anomaly decision. For example, when a spike in network latency is flagged as anomalous, feature attribution may indicate that the deviation is driven primarily by unusual memory usage rather than CPU load. Such insights are essential for root cause analysis, model validation, and debugging by domain experts (Choi et al., 2022). In addition, model-agnostic explanation methods can be integrated to provide interpretable outputs even for black-box models.

Visual interpretability concerns the presentation of anomaly detection results through intuitive and informative visual representations. Effective visualization enables system operators and analysts to assess the validity of detected anomalies, explore deviations across time and variables, and make informed decisions based on system feedback. By rendering complex statistical or algorithmic outputs accessible to non-technical stakeholders, visual interpretability significantly enhances the usability and operational integration of anomaly detection frameworks (Zhong et al., 2023).

5.6. Data quality-aware anomaly detection feature

Anomaly detection frameworks are rarely applied to fully reliable data sources; instead, they typically operate under conditions of uncertainty, noise, missing values, or partial system failure. To maintain performance and reliability under such conditions, frameworks must incorporate mechanisms that recognize and adapt to data quality issues. This motivates the inclusion of data quality–aware anomaly detection as an important feature dimension, emphasizing that effective anomaly detection must be sensitive to the integrity, completeness, and reliability of input data (Kittler et al., 2014). When properly integrated, data quality awareness improves detection precision and the trustworthiness of system outputs. This feature dimension is defined by three key subcategories.

Data quality handling refers to a framework's ability to process and remain robust to common data inconsistencies, including noise, missing values, and outliers. In many domains, such as sensor networks, log telemetry, and business analytics, data degradation is unavoidable (Kittler et al., 2014). To mitigate its impact, frameworks employ preprocessing or model-level strategies, such as imputation, smoothing, or robust statistical techniques, ensuring that detected anomalies are not artefacts of poor data quality (Alam et al., 2019). Data quality handling focuses on input resilience, and data integrity monitoring concerns a framework's ability to detect failures in the data collection, transmission, or storage pipeline. These failures may include corrupted files, communication breakdowns, timestamp misalignment, or format inconsistencies, all of which can compromise anomaly detection results. Frameworks equipped with integrity monitoring mechanisms may enforce schema validation, track metadata consistency, or generate alerts when data streams deviate from expected structural norms (Sunny et al., 2022). This capability is particularly important in automated pipelines, where detection reliability depends on the integrity of upstream data sources. Beyond technical integrity, data reliability and trustworthiness capture a framework's ability to assess and communicate confidence in both the input data and its anomaly detection outputs. This may involve assigning reliability scores to data sources, quantifying uncertainty in model predictions, or flagging results derived from potentially compromised inputs.

6. Anomaly detection frameworks' alignment with feature sSet

This research introduces the Anomaly Detection Framework Feature Set (ADF²S) to analyze the selected anomaly detection frameworks. The objective is to assess how ADF²S aligns with the architectural and functional characteristics of these frameworks. Accordingly, a systematic evaluation of Opprentice (Liu et al., 2015), the Ensemble Anomaly Detection (EAD) framework (Araya et al., 2017), and the Microsoft Time-Series Anomaly Detection service (Ren et al., 2019) is conducted using the proposed ADF²S. Table 1 summarizes the presence or absence of each feature, while the subsequent discussion provides interpretative insights into the architectural focus, strengths, and limitations of each framework. The evaluation follows a structured feature-mapping methodology, in which each ADF²S feature and sub-feature is examined against the technical descriptions, architectural details, and methodological explanations reported in the corresponding publications. For each feature, the degree of support is assessed using a three-level classification scheme: (✓) Supported, indicating that the framework explicitly implements or relies on the feature as part of its core design; (✗) Not Supported / Not Mentioned, indicating that no evidence of the feature is provided or that the capability is absent; and (!) Partially Supported, denoting cases in which a feature is indirectly present, incompletely supported, applied only in a limited context, or implied without full integration into the framework's operational pipeline. This interpretive classification enables a consistent and reproducible evaluation across heterogeneous frameworks, particularly given the diversity of algorithmic approaches, data modalities, and architectural assumptions represented in the selected works.

The Opprentice framework places strong emphasis on data preprocessing and the use of supervised machine learning for anomaly classification. It supports both univariate and multivariate time-series data and employs an ensemble of statistical detectors for feature extraction prior to classification using a Random Forest model (Liu et al., 2015). This design enables effective detection of point anomalies in well-labelled batch datasets. Opprentice also incorporates a basic self-tuning adaptation mechanism through threshold smoothing, providing limited adaptive behaviour. However, the framework does not support real-time streaming and instead relies on periodic batch retraining. In addition, Opprentice lacks model real-time update capabilities, support for contextual or collective anomaly types, and feature attribution mechanisms for interpretability. The framework also does not explicitly address data integrity monitoring or provide explainable outputs, which limits its applicability in regulated or safety-critical environments.

The Ensemble Anomaly Detection (EAD) framework is specifically tailored to the energy consumption domain and employs supervised learning through a combination of pattern-based collective contextual anomaly detection using sliding windows (CCAD-SW) and prediction-based detectors, including Support Vector Regression (SVR) and Random Forest models (Araya et al., 2017).

Table 1: Anomaly Detection Frameworks Feature Set (ADF^{2}S) Alignment

Anomaly Detection Framework Feature Set (ADF^{2}S)			Opprentice anomaly detection architecture (Liu, 2015)	Ensemble anomaly detection (EAD) framework (Araya, 2017)	Time series anomaly detection at Microsoft service (Ren, 2019)
Data Handling & Input Capabilities	Multi-Modal Data	Time-series Data	✓	✓	✓
		Graph Data	X	X	X
		Log Data	X	X	X
	Data Processing	Real-time streaming	X	X	✓
		Batch processing	✓	✓	✓
	Multivariate Support	Univariate Data	✓	✓	✓
		Multivariate Data	✓	✓	✓
Anomaly Detection Type Support	Point Anomalies		✓	✓	✓
	Contextual Anomalies		X	!(partial)	✓
	Collective Anomalies		X	✓	✓
Anomaly Detection Technique Flexibility	Statistical Methods		✓	✓	✓
	Machine Learning	Supervised Learning	✓	✓	X
		Unsupervised Learning	X	X	✓
		Semi-supervised Learning	X	X	✓
Adaptability	Concept Drift Handling		!(partial)	X	✓
	Model Real-Time Update Support		X	X	✓
	Self-Tuning Adaptation		✓	X	✓
Explainability and Intepretability	Feature Attribution		X	X	!(partial)
	Visual Interpretability		✓	X	✓
Data Quality Aware	Data Quality Handling		✓	X	✓

✓ **Supported** - The framework fully implements this feature as part of its design or methodology.
X **Not Supported / Not Mentioned** - The framework does not implement or indicate the presence of the feature.
! (Partial) - The framework provides limited or indirect support for this feature.

It supports multivariate data by integrating contextual features such as weather conditions and occupancy information, and is among the few frameworks that explicitly address collective anomalies. EAD's primary strengths lie in its high detection accuracy within its target domain and its modular ensemble-based architecture. Nevertheless, the framework is fundamentally batch-oriented and lacks support for real-time streaming, incremental model updates, and adaptive threshold tuning. Moreover, EAD does not incorporate data quality assessment mechanisms or provide explainability and interpretability features, such as analytics or feature analysis.

The Microsoft Time-Series Anomaly Detection service represents a production-grade system designed for scalability and operational efficiency (Ren et al., 2019). It is engineered to process millions of time-series metrics daily and supports unsupervised and semi-supervised detection techniques based on the integration of Spectral Residual (SR) and Convolutional Neural Networks (CNNs). The framework offers real-time streaming capabilities, adaptive learning, and concept drift handling through synthetic anomaly generation and feedback mechanisms. It also includes interpretability support in the form of dashboards, reports, and diagnostic outputs, making it well-suited for cloud service monitoring and enterprise telemetry analysis. Additionally, it does not explicitly address data integrity monitoring or data reliability and trustworthiness assessment. While visual diagnostics enhance operational transparency, the framework lacks fine-grained feature-level attribution methods characteristic of more recent explainable machine learning approaches.

Although the selected anomaly detection frameworks demonstrate advanced capabilities and have a significant impact in the field, none fully support the complete Anomaly Detection Framework Feature Set (ADF2S). Opprentice is optimized for supervised batch learning, offering moderate robustness but limited scalability and real-time adaptability. The Ensemble Anomaly Detection (EAD) framework provides high domain-specific accuracy but remains static and lacks transparency in interpretability. In contrast, Microsoft's Time-Series Anomaly Detection is the most comprehensive and scalable, though it sacrifices deep model explainability for high-throughput, unsupervised deployment.

These findings highlight the trade-offs between adaptability, interpretability, and scalability that current frameworks must navigate. They also emphasize the need for modular and extensible designs, where features such as data quality awareness, explainability, and adaptability are integral components of anomaly detection systems. Comparative analysis of Opprentice (Liu et al., 2015), The Ensemble Anomaly Detection (EAD) (Araya et al., 2017), and Microsoft's Time-Series service (Ren et al., 2019) reveals distinct design methodologies and feature priorities aligned with their respective application domains and operational contexts. This feature-based comparison also identifies gaps that future research and framework development can address.

7. Conclusions and discussion

This study proposes a structured set of features for evaluating and designing anomaly detection frameworks, offering a comprehensive taxonomy of functional and architectural dimensions that define the capabilities of such systems. The anomaly detection framework feature set (ADF^{2}S) encompasses six major dimensions: Data

Handling and Input Capabilities, Anomaly Detection Type Support, Detection Technique Flexibility, Adaptability, Explainability and Interpretability, and Data Quality-Aware Anomaly Detection, each subdivided to capture more granular operational characteristics. Through a systematic analysis of three influential frameworks, the proposed ADF^{2}S proved effective in identifying the architectural strengths and limitations of each system. Opprentice framework demonstrated strong support for multivariate, supervised learning with efficient data preprocessing and limited self-tuning capabilities, but lacked real-time adaptability, interpretability, and data integrity and reliability capabilities. The Ensemble Anomaly Detection (EAD) framework, while rich in domain-specific integration and collective anomaly support, remained static, with no support for adaptive learning and data quality-aware anomaly detection. In contrast, Microsoft Time Series Anomaly Detection emerged as the production-ready solution, offering scalability, streaming data support, unsupervised adaptability, and visual interpretability, though it lacked broader data quality support and deep feature-level explainability. These findings reveal that no single framework fully satisfies all feature sets, highlighting gaps in scalability, adaptability, and interpretability. Moreover, the absence of data quality-aware anomaly detection considerations across the framework's points to a gap in current state-of-the-art developments. The proposed anomaly detection framework feature set (ADF^{2}S), thus offers practical guidance for designing anomaly detection frameworks that are modular, adaptive, interpretable, and robust to data fluctuations.

Future research may further extend the ADF^{2}S taxonomy toward the development of a unified anomaly detection framework that combines architectural abstraction with domain-specific adaptability. Although ADF^{2}S already provides a structured and comprehensive lens for the qualitative analysis of anomaly detection frameworks, this study primarily focuses on taxonomic synthesis rather than empirical validation. Consequently, quantitative metrics demonstrating the utility or performance of ADF^{2}S are not included. Future studies could operationalise the ADF^{2}S into measurable constructs, such as feature-coverage indices, framework-completeness scores, or decision-support utility measures to empirically assess how effectively the feature set supports system design, comparison, and selection. Such quantitative extensions, aligned with approaches adopted in related domain-specific anomaly detection evaluations, are introduced in the research study (Zadeja and Osis, 2025), thereby strengthening the evidence base of ADF^{2}S and broadening its applicability across diverse anomaly detection contexts.

Acknowledgements

The author has used Grammarly (Grammarly, 2025) writing assistant platform for grammatical and paraphrasing needs. The research paper is written by the author exclusively.

References

Aggarwal, C. C. (2013). *An introduction to outlier analysis*, in *Outlier Analysis*, Springer, New York. DOI: 10.1007/978-1-4614-6396-2_1.

Ahmed, M., Mahmood, A. N., Hu, J. (2015). A survey of network anomaly detection techniques, *J. Netw. Comput. Appl.* DOI: 10.1016/j.jnca.2015.11.016.

Alam, M. R., Gerostathopoulos, I., Prehofer, C., Attanasi, A., Bures, T. (2019). A framework for tunable anomaly detection, in *Proc. IEEE Int. Conf. Software Architecture (ICSA 2019)*, Hamburg, Germany, pp. 201–210. DOI: 10.1109/ICSA.2019.00029.

Al-Faouri, A. H. (2023). Adopting data mining as a knowledge discovery tool: The influential factors from the perspectives of information systems managers, *Inf. Sci. Lett.* **12**(5), 1851–1861. DOI: 10.18576/isl/120529.

Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A., Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption, *Energy Build.* **151**, 191–206. DOI: 10.1016/j.enbuild.2017.02.05.

Barnard, J., Stryker, C. (2023). *What is anomaly detection?* Available at https://www.ibm.com/think/topics/anomaly-detection.

Bhuyan, M. H., Bhattacharyya, D. K., Kalita, J. K. (2014). Network anomaly detection: Methods, systems and tools, *IEEE Commun. Surv. Tutor.* **16**(1), 303–336. DOI: 10.1109/SURV.2013.052213.00046.

Bose, I., Mahapatra, R. K. (2001). Business data mining—a machine learning perspective, *Inf. Manage.* **39**(3), 211–225. DOI: 10.1016/S0378-7206(01)00091-X.

Cauteruccio, F., Cinelli, L., Corradini, E., Terracina, G., Ursino, D., Virgili, L., Fortino, G. (2021). A framework for anomaly detection and classification in multiple IoT scenarios, *Future Gener. Comput. Syst.* DOI: 10.1016/j.future.2020.08.010.

Chai, Z., Liu, Y., Wang, M., Guo, Y., Shi, F., Li, Z., Du, J. (2022). Quantum anomaly detection of audio samples with a spin processor in diamond, preprint, arXiv:2201.10263. DOI: 10.48550/arXiv.2201.10263.

Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: A survey, *ACM Comput. Surv.* **41**(3), 1–58. DOI: 10.1145/1541880.1541882.

Chandola, V., Banerjee, A., Kumar, V. (2012). Anomaly detection for discrete sequences: A survey, *IEEE Trans. Knowl. Data Eng.* **24**, 823–839. DOI: 10.1109/TKDE.2010.235.

Chatterjee, A., Ahmed, B. S. (2022). IoT anomaly detection methods and applications: A survey, *Internet Things*. DOI: 10.1016/j.iot.2022.100568.

Choi, H., Kim, D., Kim, J., Kim, J., Kang, P. (2022). Explainable anomaly detection framework for predictive maintenance in manufacturing systems, *Appl. Soft Comput.* DOI: 10.1016/j.asoc.2022.109147.

Choi, K., Yi, J., Park, C., Yoon, S. (2021). Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines, *IEEE Access* **9**, 120043–120065. DOI: 10.1109/ACCESS.2021.3107975.

Davenport, T. H. (2005). *Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers*, Harvard Business Review Press, Boston.

Eberle, W., Holder, L. B. (2015). Scalable anomaly detection in graphs, *Intell. Data Anal.* **19**, 57–74. DOI: 10.3233/IDA-140696.

Fan, Z., You, Z. (2024). Research on network intrusion detection based on XGBoost algorithm and multiple machine learning algorithms, in *Proc. Int. Conf.*, pp. 161–166. DOI: 10.54254/2753-8818/31/20241171.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework, in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining*, pp. 82–88.

Feng, W., Cao, Y., Chen, Y., Wang, Y., Hu, N., Jia, Y., Gu, Z. (2025). Multi-granularity user anomalous behavior detection, *Appl. Sci.* **15**(1). DOI: 10.3390/app15010128.

Filzmoser, P., Maronna, R., Werner, M. (2008). Outlier identification in high dimensions, *Comput. Stat. Data Anal.* **52**(3), 1694–1711. DOI: 10.1016/j.csda.2007.05.018.

Frawley, W. J., Piatetsky-Shapiro, G., Matheus, C. J. (1992). Knowledge discovery in databases: An overview, *AI Mag.* **13**(3). DOI: 10.1609/aimag.v13i3.1011.

Grammarly (2025). *Grammarly*. Available at https://www.grammarly.com.

Goldstein, M., Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data, *PLoS ONE* **11**(4). DOI: 10.1371/journal.pone.0152173.

Herr, D., Obert, B., Rosenkranz, M. (2021). Anomaly detection with variational quantum generative adversarial networks, *Quantum Sci. Technol.* DOI: 10.1088/2058-9565/ac0d4d.

Injadat, M., Salo, F., Nassif, A. B., Essex, A., Shami, A. (2018). Bayesian optimization with machine learning algorithms towards anomaly detection, in *Proc. IEEE GLOBECOM 2018*, Abu Dhabi, United Arab Emirates, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647714.

Jansevskis, M., Osis, K. (2024). Securing the future: The role of knowledge discovery frameworks, in Sipola, T., Alatalo, J., Wolfmayr, M., Kokkonen, T. (eds), *Artificial Intelligence for Security*, Springer, Cham. DOI: 10.1007/978-3-031-57452-8_5.

Kittler, J., Christmas, W., de Campos, T., Windridge, D., Yan, F., Illingworth, J., Osman, M. (2014). Domain anomaly detection in machine perception: A system architecture and taxonomy, *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(5), 845–859. DOI: 10.1109/TPAMI.2013.209.

Klosgen, W., Zytkow, J. (1996). Knowledge discovery in databases terminology, in Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds), *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge, MA, pp. 573–592.

Liu, D., Zhao, Y., Xu, H., Sun, Y., Pei, D., Luo, J., Feng, M. (2015). Opprentice: Towards practical and automatic anomaly detection through machine learning, in *Proc. ACM SIGCOMM Internet Measurement Conference (IMC 2015)*, Tokyo, Japan. DOI: 10.1145/2815675.2815679.

Mahapatra, I., Bose, R. K. (2001). Business data mining—a machine learning perspective, *Inf. Manage.*, 211–225. DOI: 10.1016/S0378-7206(01)00091-X.

Marsland, S. (2001). *On-line novelty detection through self-organisation, with application to inspection robotics*, PhD thesis, University of Manchester.

Mishra, S., Kumar, M. (2022). CNN-based anomaly detection in complex systems, *Computology: J. Appl. Comput. Sci. Intell. Technol.*, 54–57. DOI: 10.17492/computology.v2i2.2206.

Nassif, A., Talib, M., Nasir, Q., Dakalbab, F. (2021). Machine learning for anomaly detection: A systematic review, *IEEE Access* **9**, 78658–78700. DOI: 10.1109/ACCESS.2021.3083060.

Nonaka, I., Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, Oxford. DOI: 10.1093/oso/9780195092691.001.0001.

Pang, G., Shen, C., Cao, L., van den Hengel, A. (2021). Deep learning for anomaly detection: A review, *ACM Comput. Surv.* **54**(2). DOI: 10.1145/3439950.

Pimentel, M. A., Clifton, D. A., Clifton, L., Tarassenko, L. (2014). A review of novelty detection, *Signal Process.* **99**, 215–249. DOI: 10.1016/j.sigpro.2013.12.026.

Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Zhang, Q. (2019). Time-series anomaly detection service at Microsoft, in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. DOI: 10.48550/arXiv.1906.03821.

Steenwinckel, B., Paepe, D., Hautte, S., Heyvaert, P., Bentefrit, M., Moens, P., Van Herwegen, S. (2021). FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing learning, *Future Gener. Comput. Syst.*, 30–48. DOI: 10.1016/j.future.2020.10.015.

Sunny, J. S., Patro, C. P., Karnani, K., Pingle, S. C., Lin, F., Anekoji, M., Ashili, S. (2022). Anomaly detection framework for wearables data: A perspective review on data concepts, data analysis algorithms and prospects, *Sensors* **22**(3). DOI: 10.3390/s22030756.

Thudumu, S., Branch, P., Jin, J., Singh, J. (2020). A comprehensive survey of anomaly detection techniques for high-dimensional big data, *J. Big Data*. DOI: 10.1186/s40537-020-00320-x.

Torr, P., Murray, D. (1993). Outlier detection and motion segmentation, in *Proc. SPIE*, pp. 432–443.

Wang, F., Jiang, Y., Zhang, R., Wei, A., Xie, J., Pang, X. (2025). A survey of deep anomaly detection in multivariate time series: Taxonomy, applications, and directions, *Sensors*. DOI: 10.3390/s25010190.

Xu, H., Pang, G., Wang, Y., Wang, Y. (2023). Deep isolation forest for anomaly detection, *IEEE Trans. Knowl. Data Eng.* **35**(12), 12591–12604. DOI: 10.1109/TKDE.2023.3270293.

Zadeja, I., Osis, K. (2025). Anomaly detection framework in cybersecurity: Features and design considerations, in Wolfmayr, M., Sipola, T., Kokkonen, T. (eds), *Emerging Technologies in Cybersecurity*, Springer.

Zhong, L., Zhang, Y., van Leeuwen, M. (2023). A survey on explainable anomaly detection, *ACM Trans. Knowl. Discov. Data* **18**(1). DOI: 10.1145/3609333.

Zhou, Y., Ren, H., Li, Z., Pedrycz, W. (2022). Anomaly detection based on a granular Markov model, *Expert Syst. Appl.* **187**. DOI: 10.1016/j.eswa.2021.115744.

Zuo, X., Yang, X., Dou, Z., Wen, J. (2019). Experience report: Deep learning-based system log analysis, *TREC Proc.* DOI: 10.48550/arXiv.2107.05908.

Received May 24, 2025, revised December 8, 2025, accepted January 16, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 70–93
https://doi.org/10.22364/bjmc.2026.14.1.04

Digital Competencies and Barriers of Technology Adoption in Care Giving: A Systematic Literature Review

Palmi LAHE, Tarmo ROBAL

Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia
palmi.lahe@taltech.ee, tarmo.robal@taltech.ee

ORCID 0009-0002-3702-0294, ORCID 0000-0002-7396-8843

Abstract. The ageing population, rising healthcare costs and shortage of professional caregivers are making significant changes to health and care systems around the world, including an increased demand for services, higher costs, workforce shortages, and a greater reliance on technology like tele-health and AI to improve efficiency and access to care. Therefore, having the competence and ability to use digital tools is compelling for professional caregivers and policymakers alike, as it underpins successful technology adoption and digital transformation. Yet, today, professional caregivers lack knowledge and skills that would support technology adoption and digital transformation. Most countries have set high ambitions to strengthen digitalisation of their citizens and society. This review article provides an overview of the challenges faced by the welfare sector, professional caregivers, and care homes due to the lack of or inadequate digital competencies among workers, highlighting how this has hindered the adoption of technology. Through a systematic literature review, 59 studies from the last 10 years were identified and obtained from well-known databases. The review of these papers identifies the gaps: there is often insufficient training for staff on how to use and integrate assistive technologies effectively, or staff may be unaware of the available technologies. Challenges such as technological limitations and low involvement of professional caregivers are noted. The literature review suggests that for technology adoption in the welfare and caregiving sector, especially for emerging technologies, it is necessary to prioritise and recognise professional caregivers' digital competencies, and find solutions for existing barriers.

Keywords: digital competencies, professional caregiver, emerging technology, welfare technology, barriers for technology adoption

1 Introduction

As we progress further into the 21st century, an unprecedented rapid increase in the total and relative number of older persons in both the developed and the developing world is being observed. The World Health Organization is tackling the multifaceted challenges of population ageing through a comprehensive approach that includes consolidating databases, building capacity through research and training, advocating for change, and developing effective policies (De Carvalho et al., 2017; OECD Health Policy Studies, 2020). Population ageing is leading to an increase in the prevalence of certain diseases, while advancements in technology and digital health monitoring tools (Zyzak and Martinussen, 2024) are driving the future need for professional caregivers. Additionally, socio-economic developments such as ageing populations, changing family structures, and rising economic pressures are leading to a decline in care, which is expected to increase the demand for welfare services in the coming years (European Commission, 2021; OECD Health Policy Studies, 2020).

Welfare technology, a concept originating from Scandinavia from the National Board of Health and Welfare in Sweden, includes digital tools aimed at maintaining and increasing safety, activity, participation, or independence for individuals at risk of impairment (Zander et al., 2021). The term welfare technology is broad and applicable to various technologies improving life quality, such as electronic medicine tablet dispensers, reminders, safety alarms, trackers, bed sensors, virtual reality (VR) for cognitive stimulation, smart door locks for keyless access, humanoid and pet robots, telepresence, rehabilitation, entertainment, domestic and fall detection/prevention, or social robots (Spann et al., 2022). Other commonly used terms are, emerging, assistive living technology and e-health (Zander et al., 2021).

professional caregivers in the welfare sector are engaged with personal care and assistance of patients and elderly, convalescent and disabled people in health care and residential settings. Jobs within this sector include healthcare assistants and home-based personal professional caregivers (Elfert, 2019). In this literature review, we focus on professional caregivers who work with the elderly in care homes or in home care settings. For the elderly, welfare technologies can be life changers in helping them to live more independently, stay connected with the world, and manage their health effectively (Zander et al., 2021). Yet, to be able to manage such technologies, they need support and help from competent professional caregivers. Therefore, the problem is twofold – for successful welfare technology adoption support is needed for those who benefit, and the supporting personnel, the professional caregivers, require digital competencies to be on top of such technologies.

Digital competence is the general term used to describe or explain the ability (of a citizen, a student, a teacher, etc.) to use information technology (IT) in a specific contex (Rizza, 2014). Digital competencies are increasingly essential as technology becomes integral to welfare in the daily support for elderly care and care of vulnerable individuals. Digital competence refers to the confident, critical, and responsible use of digital technologies for learning, work, and participation in society (Vuorikari et al., 2022). It encompasses a combination of knowledge, skills, and attitudes necessary to effectively navigate and utilize digital environments. Digital competence development is increasingly framed as a lifelong learning process, particularly for adults whose pro-

fessional roles evolve alongside rapid technological change (Platonova et al., 2024). European Council (European Commission, 2021; Kotsiou et al., 2022), has listed digital competence as one of the eight key competencies that support effective participation in education, work, and personal lives in present and future societies and economies. The DigComp framework (Vuorikari et al., 2022), valid throughout Europe, has been created to assess the digital competence of citizens in the EU. The DigComp framework creates an agreed vision in terms of competencies, which helps to reach a common understanding and solve the challenges that arise almost everywhere with digitalization (Sulakatko, 2024). It also acts as a guide for educational institutions and learners on their digital competence. Educational institutions, such as vocational schools for professional caregivers, can align their curricula, training activities, and programs for lifelong learning with the guidance of DigComp framework. In addition, the framework enables to evaluate individual digital competence. However, a significant segment of the framework has been missed from the DigComp framework of competencies. Specific digital competencies related to health and safety, such as using technology to monitor patients' health or manage emergencies, could be more explicitly detailed. Furthermore, DigComp framework does not include field-specific digital competencies, highlighting the need to identify IT skills specifically required by professional caregivers to use digital tools in their profession.

Digital competence is not the only factor influencing technology adoption. Technology-related impediments comprise factors such as technology anxiety and resistance to technology (Talwar et al., 2024), lack of technical competence (Spånt Enbuske, 2022), and lack of ability to operate equipment, low computer self-efficacy and fears of disruption of services (Nilsson et al., 2020). Resistance factors such as uncertainty, security concerns, privacy issues, lack of trust, threats to identity, information overload, system feature overload, fatigue, and technostress, can be exacerbated by insufficient training (Zander et al., 2021). It has been shown that anxiety among healthcare staff regarding the delivery of proper patient care with new technology, driven by self-doubt about their technical competence, skills, and knowledge, acts as a significant dissuading factor (Andersson et al., 2017). Several studies (Gullslett et al., 2022; Blindheim et al., 2022; Glomsås et al., 2020), have noted the impact of technical complexities and technical capabilities as reasons for resistance towards technological innovations. This leads us to a statement that professional caregivers need to be digitally fluent and stay on top of the latest care innovations for technology adoption readiness and new roles digitalisation and new digital tools introduce.

The aim of this review article is to identify the challenges faced by professional caregivers due to the lack of digital competencies. To synthesize and summarize the literature regarding the challenges and barriers of implementing emerging welfare technology, policy, and practice the following research questions are used:

RQ1: What are the key digital competencies acting as enablers for professional caregivers to effectively implement emerging technologies in the welfare sector?

RQ2: What barriers exist for professional caregivers in adopting and using welfare technologies and digital platforms?

The rest of the paper is organized as follows: Section 2 discusses the theoretical background, Section 3 describes the research method, Section 4 describes the results, Section 5 provides discussion, and Section 6 is summing up this literature review.

2 Theoretical Background

The transition to digitalization and the larger process of digital transformation began during the economic crisis of 2008-2009 (Novichenko, 2020). The latter is considered crucial for social development and progress, emphasizing the ongoing need for individuals to acquire and enhance digital skills and competencies. The EU's approach focuses on creating conditions that enable citizens to effectively utilize digital technologies while safeguarding themselves against potential risks in the digital environment. Additionally, it is vital to continually update these skills and recognize their importance in navigating an increasingly digital world (Vuorikari et al., 2022). Digitally competent citizens are the foundation of the digital state and a key element in driving economic development. Continuous digital transformation of countries has been regarded as one of the crucial steps in overcoming economic downturns (Novichenko, 2020). The future skills required for professional caregivers are shaped by the specific demands of the care sector as well as broader structural factors. According to reports by the European Commission and the European Centre for the Development of Vocational Training (Cedefop, 2024), several key factors will shape the future demand for care and the necessary skill sets for professional caregivers. One significant factor is the increasing integration of technology into healthcare, which is expected to improve access to high-quality, affordable care services and support independent living (European Commission, 2021). However, this technological advancement will require professional caregivers to undergo additional training.

The key factors shaping the future demand for professional caregivers include technical proficiency, integration skills, ethical competence, and a patient-centred approach (Schwiter and Steiner, 2020). Digital devices are transforming care work by enabling remote tasks and precise health monitoring. To thrive in this evolving landscape, professional caregivers must develop essential cross-disciplinary skills such as digital competence, problem-solving, initiative, continuous learning, cultural awareness, resilience, social intelligence, creativity, critical thinking, and adaptability (Schwiter and Steiner, 2020). Education and training play a vital role in fostering digital competence by providing essential activities that enhance knowledge, skills, and expertise (Cedefop, 2024).

Continuous education and training are crucial for professional caregivers, it enables to update their skills and knowledge, ensure adaptation to new technologies, and provide high-quality care (Lahe, 2025). Lifelong learning, as defined by the EU (European Commission, 2021), encompasses all learning activities throughout life aimed at improving knowledge, skills, and competencies in personal, civic, social, or employment-related contexts. For professional caregivers, digital competence involves both technical skills and digital literacy (Anthonysamy et al., 2020). Technical skills include operating and troubleshooting welfare monitoring tools, while digital literacy encompasses the ability to access, manage, evaluate, and create information. These skills help pro-

fessional caregivers to navigate digital environments safely and effectively, enhancing their ability to deliver personalized and efficient care (Rosell Magerøy et al., 2023).

3 Research Method

A systematic literature review was conducted to compile and analyse the key digital competencies that enable professional caregivers to effectively implement emerging technologies, as well as the barriers existing for the implementation in the welfare sector. Systematic literature review is a form of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and repeatable (Kitchenham et al., 2009). In this literature review, the Kitchenham method, which provides structured guidelines for defining search strategy allowing to choose studies that are closely related to the topic, was applied.

3.1 Search Strategy

After formulating the research questions, a targeted search strategy was developed to identify studies that provide evidence on the challenges and gaps in digital competencies, as well as the barriers professional caregivers encounter in implementing emerging technologies in the welfare sector. A set of keywords was identified as follows: *professional caregiver*, *digital competence*, *barriers* and *digital platforms*, and combined with *welfare technology* as described in Table 1. These keywords were identified based on the set research questions and expanded based on the results of quick pilot searches in key databases by inspecting titles and abstracts.

The selected terms were combined to search strings to ensure that all significant research on the utilization of digital competence and main gaps of technology usage for professional caregivers is included. Two search strings were composed targeting either RQ1 or RQ2. Using two different search strings helps make the literature search more complete and balanced. Each search string may capture studies that the other might miss, so using both increases the chances of finding all relevant research. When the results from both searches are combined, the review can address both research questions more effectively. This approach ensures that any overlapping or complementary studies are included and considered for both questions when appropriate. Overall, it leads to a fuller and more accurate understanding of the topic by drawing on the strengths of both search strategies.

The scope of the review was set to specifically target the following types of papers: study, literature review, ethnographical study, comparative study, scoping review, and pilot study due to their relevance and contribution to advancing knowledge in the field. The studies were limited to those published in English language with accessibility of findings for an international audience, and published within the time frame of last 10 years. The search was conducted against key academic resources such as IEEE Explore, ACM Digital Library, ScienceDirect, Wiley Online Library, PubMed, SpringerLink, and Taylor and Francis. Table 2 provides the details of the search.

Table 1. Description of used keywords.

Keyword	Description	Justification
professional caregiver	Individuals who provide care, often in a professional or familial role of patients, elderly or disabled people in health care and residential	Welfare technology is designed to support professional caregivers, easing their tasks and improving the quality of care they provide.
Digital Competence	Skills and knowledge necessary to effectively use digital tools and technologies	professional caregivers must possess digital competence to efficiently utilize welfare technologies, improving the quality of caregiving.
Barriers	Challenges that prevent or hinder the adoption or use of welfare technology	Identifying barriers, such as lack of training or resources, is essential to overcome obstacles in the successful use of welfare technologies.
Digital Platforms	Digital tools, systems, and interfaces used to manage and facilitate care giving tasks	Digital platforms play a critical role by enabling professional caregivers to access information, communicate and manage caregiving responsibilities more effectively.
Welfare Technology	Welfare technology refers to the digitalisation of health care and social services	For the elderly, welfare technologies can be life changers in helping them to live more independently, stay connected with the world, and manage their health.

Table 2. Details of the search.

Electronic databases	IEEE Xplore, ACM Digital Library, Scopus, ScienceDirect, Wiley online library, PubMed, SpringerLink, Taylor & Francis journals
Search strings	RQ1: (professional caregiver OR welfare) AND ”digital competence” AND ”welfare technology” RQ2: (professional caregiver OR welfare) AND ”welfare technology” OR ”digital platform” AND barrier
Type of searched literature	Study, Literature Review, Ethnographical Study, Comparative Study, Scoping Review and Pilot Study
Language of study	English
Publication period	From October 2014 to November 2024

3.2 Inclusion and Exclusion Criteria

On top of the search scope of the review (Section 3.1), additional inclusion and exclusion criteria were established. This refinement enhances the quality and focus of the review by including only studies on welfare technology that specifically address the key aspects under investigation. Following the principles of the Kitchenham (Kitchenham et al., 2009) method, these criteria were defined to ensure transparency, reproducibility, and relevance of the selected studies. In line with Kitchenham's systematic review guidelines, the inclusion and exclusion process aimed to minimize bias and maintain a clear link between the research questions and the reviewed evidence. The inclusion criteria were:

- Studies were included if they involved key stakeholders in elderly care, such as care workers, professional caregivers, managers, the elderly, care homes, home care services, or governmental bodies responsible for policies;
- The study had to focus on digital competencies, welfare technology in relation to professional caregivers, specifically, studies addressing barriers, training, competencies, or gaps in the use or implementation of digital and welfare technologies were considered relevant;
- The study was required to be long-term, lasting at least a month, to promote sustained engagement and meaningful learning outcomes. Studies with a duration of less than one month were excluded, as shorter studies may not provide sufficient time to observe meaningful outcomes, behavioral changes, or sustained effects of welfare technology interventions.

Additionally, exclusion criteria were integrated to strengthen the process of searching for evidence when assessing the relevance of each screened article:

- Studies not involving professional caregivers were excluded;
- Studies not focused on elderly care were excluded;
- Studies with a duration of less than one month were excluded.

These criteria ensured a systematic and rigorous selection process, focusing on high-quality and relevant studies for the literature review.

3.3 Study Search and Selection

The initial search carried out between 09–10 October 2024 identified 505 articles as follows: IEEE Xplore (n=14), The ACM Digital Library (n=183), ScienceDirect (n=82), PubMed (n=13), Scopus (n=13), Tandfonline (n=94), SpringerLink (n=59), and Wiley Online (n=47). During the identification step, the duplicates were removed (n=82), leaving 423 records for screening.

Next, during screening, the titles and abstracts of the papers were analysed, and papers not falling within the scope of this research, i.e., not addressing professional caregivers or elderly care in the research context (Section 3.2) (n=310) were excluded, reducing the number of articles to 113 for full-text review during eligibility check. During full-text analysis, of these 54 records were found not to discuss professional caregivers in elderly care context, resulting in their removal. The final study set therefore

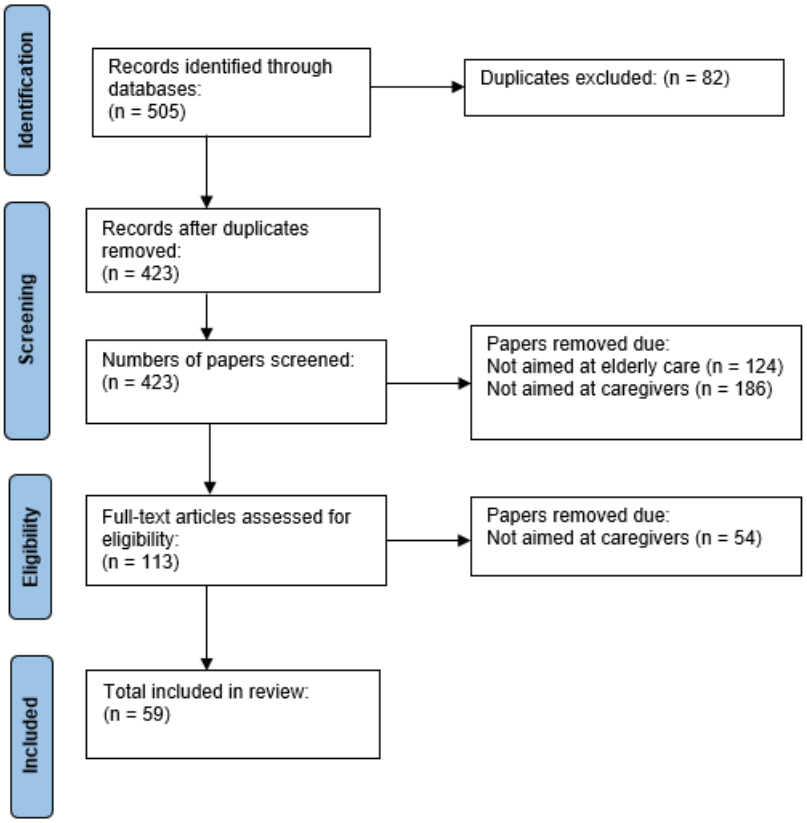


Fig. 1. The PRISMA flow diagram for this study adapted from (Haddaway et al., 2022).

consists of 59 papers. This literature selection process, is illustrated in Figure 1 as a PRISMA diagram.

In the analytical phase of the review, the selected papers underwent a rigorous open coding process conducted by a single researcher. The primary aim was a comprehensive examination of the reported digital competencies required for professional caregivers (RQ1), and the barriers that hinder the effective integration of new technologies (RQ2) in the welfare sector for professional caregivers. All studies were meticulously analysed using Microsoft Excel as the analysis tool.

During the analysis, each paper was carefully read and the following aspects documented in the allocated Excel spreadsheet:

– Problems addressed in the paper (as identified by the authors, e.g., research questions). Open coding was applied to categorize the problems as: training and competence, legal and ethical issue, economic and budgetary, technological barriers and professional caregiver's input;

- Type of welfare technology used in the studies. Open coding was applied over five different types of technologies used across the studies: digital care technology, multimedia, monitoring technology, service robots, social robots;
- Technology and digital competencies demonstrated or required in the studies (as reported by the authors). Open coding was applied to classify competencies such as basic digital skills, technological adaptability, information management, and data security awareness in digital practice. These categories correspond largely to Dimension 1 of the DigComp framework (Vuorikari et al., 2022), encompassing areas such as information and data literacy, safety, digital content creation, and problem solving. On a more detailed level, they align with Dimension 2 elements including managing digital content, protecting personal data, and identifying technological needs, illustrating a close connection between the empirical findings and the established DigComp framework.
- Location of each study, identified with the granularity of country in which the research was conducted. This was included as the geographical context can significantly influence the findings due to differences in cultural norms, social structures, healthcare systems, and policy frameworks. Open coding included the names of countries, in total counting for 15 labels.
- Type of research reported in the paper. Open coding was done through the following labels: study, literature review, ethnographical study, pilot study, comprehensive review, scoping review.
- Sample size and study duration as reported in the paper.
- Research method used in the paper, through the following labels: qualitative (e.g., interviews or observations), quantitative (e.g., surveys or statistical data), or mixed methods as the combination of the two latter.
- Summary of the key findings and insights, highlighting implications for practice, policy, or further research, and often suggesting recommendations for improving the implementation or use of welfare technology.

This detailed and structured approach allowed for a thorough analysis of each article, ensuring that all relevant aspects were considered and documented. The use of Excel as a tool facilitated a structured comparison of data across different studies, providing a clear and comprehensive overview of the findings. The focus was on identifying the digital competencies mentioned in the papers (RQ1), and the gaps in the adoption and usage of emerging technologies (RQ2).

4 Results

This section presents the findings of the systematic literature review, which analyses studies conducted between October 2014 to November 2024.

We start by characterising the 59 papers that this literature review on digital competencies and barriers of technology adoption in caregiving is based on. First, to characterise the papers, we analyse the geographical scope and research design of the final set of papers.

4.1 Overview of the Studied Literature

The global significance and interest in welfare technology were demonstrated across 59 studies conducted in 16 distinct countries. Four studies addressed multiple countrie resulting in a total of 84 country contributions as illustrated in Figure 2.

Overall, Norway emerged as a leading contributor with 25 contributions (29.8%), followed by Sweden with 13 contributions (15.5%) and Finland with 12 contributions (14.3%). China contributed six studies (7.1%), while Bangladesh, Pakistan, and the United States each contributed four (4.8%). Denmark and Taiwan contributed two and three studies respectively, and Germany, South Korea, the United Kingdom, Brazil, France, Switzerland, and Spain contributed between one and two studies each.

Interestingly, welfare technologies in the context of elderly care were predominantly studied in Scandinavia. Norway, Sweden, and Finland—neighbouring countries of Scandinavia—accounted for 50 contributions (59.5%), highlighting the regional importance of welfare technologies as solutions to the ageing population challenge.

Concerning the research design, the following types could be distinguished as described in Table 3. The majority of research (72.9%) in this literature review is original research covering various welfare topics. Literature reviews account for (20.3%), offering comprehensive analyses of existing studies. Ethnographical studies, pilot studies, comparative studies, and scoping reviews each constitute a smaller share, ranging from 1.7% highlighting their more specialized roles in research. Out of the total 59 articles included in the sample, 11 (19%) employed quantitative methods, 29 (49%) used qualitative approaches, and 19 (32%) followed a mixed methods design for research.

In conclusion, the predominance of original and qualitative studies suggests that research in this field is still in an exploratory phase, focusing on understanding complex

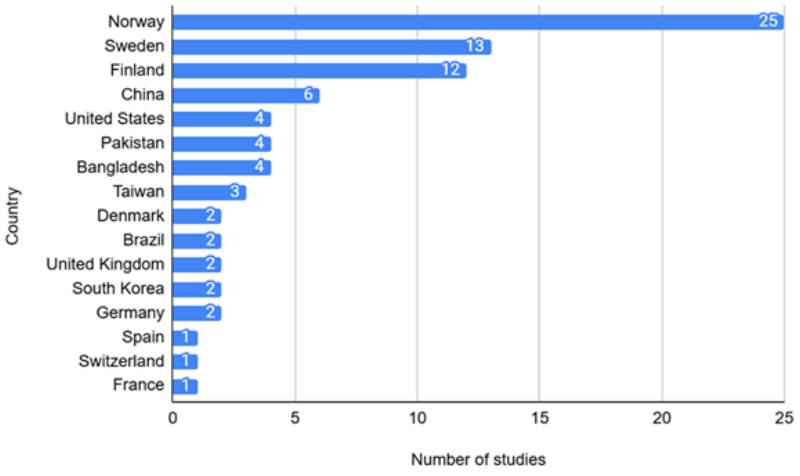


Fig. 2. Number of studies per country reported in the studied literature.

Table 3. Research Design Categories and Descriptions

Research Design Category	Description	# Total	%
Original research	Original research across various topics.	43	(72.9%)
Literature Review	Comprehensive review of existing literature.	12	(20.3%)
Ethnographical Study	In-depth qualitative analysis of cultural context.	1	(1.7%)
Pilot Study	Preliminary study to assess feasibility.	1	(1.7%)
Comparative Study	Comparison of different variables of population.	1	(1.7%)
Scoping Review	Overview of key concepts and gaps.	1	(1.7%)
TOTAL		59	100%

social, organisational, and individual factors influencing welfare and digital transformation.

4.2 Technology Usage Among professional caregivers

To better understand the types of skills that may be needed in the welfare sector by professional caregivers and answer RQ1, we examine the technologies mentioned in the reviewed papers, as summarized in Table 4.

Digital care technologies (8 studies; 14.3%), particularly e-health systems, are designed to integrate client records and improve professional caregivers' access to essential information. Studies highlight their potential, but also point to adoption challenges such as system complexity, lack of seamless integration, and user resistance (Oh and Bae, 2024; Boyle et al., 2022; Ebrahimi et al., 2023; Dugstad et al., 2019). While these tools encourage self-monitoring and patient autonomy, their success depends on professional caregivers overcoming barriers related to digital skills, workload, and technology acceptance.

Multimedia tools (8 studies; 14.3%), including virtual reality and gaming, have emerged as innovative options for rehabilitation and mental health therapy. They offer engaging ways to support cognitive and emotional well-being, especially for people with dementia (White et al., 2023; Jawharieh et al., 2024). However, their implementation remains in an early stage, and further research is needed to assess long-term impacts.

Monitoring technologies: (23 studies; 41.1%)such as alarms, cameras, door sensors, GPS-enabled devices, and digital medication dispensers—play an increasingly important role in safety and health monitoring (Flakk Nordang and Halvorsen, 2022; Sten-Gahmberg et al., 2024; Dugstad et al., 2019; Kleiven et al., 2020). While they are effective in ensuring security, studies suggest that less emphasis has been placed on improving quality of life, social participation, and independence (Glomsås et al., 2022; Gullslett et al., 2022).

Service robots: (9 studies; 16.1%) are gradually being integrated into care settings, assisting with mobility, daily tasks, and nighttime support for people with dementia (Tuisku et al., 2022; Dugstad et al., 2019). They can also support self-diagnosis and introduce novel care practices that promote independence (Pekkarinen and Melkas, 2019; Tuisku et al., 2022). Despite their potential to improve efficiency, their broader adoption remains limited by cost and technological constraints.

Social robots: (4 studies; 7.1%), such as Pepper, Zora, AIBO, and Justo Cat, are designed to provide companionship, reduce loneliness, and support physical care (Isaksen and Rypestøl, 2022; Melkas et al., 2020; Budak et al., 2021). They have been adopted for therapy, exercise, and recreational activities (Turja et al., 2022), though their acceptance varies across care settings. Some studies also note that anticipated time savings for staff have not materialized in practice (Blindheim et al., 2022).

Table 4. Technology usage among professional caregivers

Type	Details	# of studies (n)	Percentage (%)
Digital care technology	E-health databases and other e-health technologies to lessen pressure on healthcare systems.	8	14.3
Multimedia	Use of virtual reality and digital games as tools for supporting elderly cognitive and emotional well-being.	8	14.3
Monitoring technology	Sensors and telemedicine used to gather information about the environment and monitor well-being.	23	41.1
Service robots	Robotic technologies assisting professional caregivers with mobility and daily support tasks.	9	16.1
Social robots	Robots providing physical interaction, companionship, and support, and reducing loneliness.	4	7.1
Telepresence technology	Technologies enabling remote presence and digital communication in caregiving contexts.	4	7.1
Total		**59**	**100%**

Finally, telepresence technologies: (4 studies; 7.1%) enable professional caregivers to remain virtually present with clients, regardless of location (Berg et al., 2024; Boyle et al., 2022; Gates et al., 2019; Turja et al., 2022). These devices help reduce loneliness and strengthen communication, though they require stable connectivity and frequent interaction. Compared to simple video calls, embodied telepresence robots can foster more engaging and meaningful interactions.

The primary purpose of using technology in care homes is to enhance the quality of care, promote independence, and improve residents' overall well-being while supporting staff in their daily responsibilities. Service robots contribute significantly by assisting with mobility, facilitating daily routines, and providing nighttime monitoring and self-diagnostic functions. These features not only reduce staff workload but also empower residents to manage aspects of their own care, thereby fostering greater autonomy and safety. Social robots, on the other hand, play a crucial role in addressing emotional and psychological welfare. By offering companionship, engaging residents in therapeutic and recreational activities, and stimulating cognitive and physical exercise, they help reduce loneliness and enhance social interaction. Telepresence technologies further strengthen emotional and social well-being by enabling professional caregivers and family members to maintain a virtual presence, ensuring continuous communication, and mitigating social isolation—particularly for residents with limited mobility or during staff shortages. Collectively, these technologies enhance the efficiency and quality of care delivery, support residents' physical, cognitive, and emotional health, and help address workforce challenges by supplementing human labor with reliable, responsive systems.

4.3 Required digital competences for professional caregivers

The integration of emerging technologies in the welfare sector is rapidly transforming caregiving practices. These technologies, including e-health solutions, care robots, and digital platforms, hold great potential for improving care outcomes, enhancing efficiency and supporting professional caregivers in their roles. However, the effective implementation of these technologies hinges on the digital competence of professional caregivers. In order to answer RQ1 we identify and discuss the key digital competencies that according to literature enable professional caregivers to implement emerging technologies with the welfare sector.

Digital competence encompasses the knowledge, skills, and attitudes required to successfully engage with and utilize welfare technologies in day-to-day care activities, requiring professional caregivers to possess a broad range of digital competencies (Fjeldheim et al., 2024). Staff must be able to operate service robots, social robots, and telepresence devices, manage connectivity, and troubleshoot technical issues to ensure safety and continuity of care. Strong digital communication skills are also essential to facilitate virtual interactions and guide residents in the use of technology. Awareness of data privacy and ethical considerations is crucial when handling sensitive resident information. Moreover, professional caregivers must be adaptable and committed to continuous learning as technologies evolve, addressing skill gaps that could hinder effective use. Finally, they should be able to integrate digital tools meaningfully into daily care routines, enhancing residents' well-being while complementing human care.

At the same time, the ethics of care is rarely used as a theoretical framework in empirical studies of technology-mediated healthcare practices. Applying an ethics of care perspective in such contexts highlights new dilemmas, relational tensions and vulnerabilities, and requires decision-making that is informed by a higher awareness of business ethics, e-ethics, netiquette and data protection (Ramvi et al., 2023; Sulakatko, 2024). Skills-forecasting tools further support workers and learners in making informed

decisions regarding education, training, and career development opportunities (Cedefop, 2024). Moreover, significant differences persist in how digital competencies are conceptualized and developed across regions, such as between the EU (Novichenko, 2020).

Key factors include recognizing the role of digital technology in both professional and personal contexts and fostering openness to experimentation and change. Building adaptability, encouraging positive attitudes, and promoting digital literacy help reduce fear and resistance toward new technologies. Enthusiasm, commitment, and an understanding of the benefits of digital transformation are essential for enabling professional caregivers to effectively adopt and implement emerging technologies within the welfare sector (Gustafsson and Dannapfel, 2025; Fjeldheim et al., 2024).

One of the foundational skills highlighted in several studies is is basic digital literacy (Bae et al., 2024; Zander et al., 2021; Frennert, 2018; Anthonysamy et al., 2020),

Table 5. Required digital competences for professional caregivers.

Digital Areas	Competence Focus	Author(s)
Basic Digital Literacy	Foundational IT skills for professional caregivers, including the ability to use digital devices and navigate care management systems, patient information, and communication tools.	Anthonysamy et al. (2020); Bae et al. (2024); Frennert (2018); Zander et al. (2021)
Adaptability to Technology	professional caregivers' ability to learn, adapt, and troubleshoot new digital tools and systems amidst rapid technological changes in caregiving.	Eriksson and Ineland (2023); Sten-Gahmberg et al. (2024)
Data Security and Privacy Competence	Foundational IT skills for professional caregivers, including the ability to use digital devices and navigate care management systems, patient information, and communication tools.	Johansson-Pajala and Gustafsson (2020)
Effective Communication	professional caregivers' ability to use digital platforms for effective communication, including video calls, messaging systems, and collaboration tools to coordinate care.	Eriksson and Ineland (2023); Gustafsson and Dannapfel (2025); Sten-Gahmberg et al. (2024)

which refers to the ability to use digital devices (such as computers, smartphones, and tablets) and navigate various software and applications. This competence is critical for professional caregivers to interact with care management systems, access patient information, and utilize digital communication tools. Another important competence found from studies is technological adaptability, which refers to a professional caregiver's ability to learn, adjust to and troubleshoot new digital tools and systems. Given the rapid

pace of technological advancements in the welfare sector, professional caregivers need to be agile in adapting to new technologies and resolving any issues that may arise during their use (Eriksson and Ineland, 2023; Sten-Gahmberg et al., 2024). As professional caregivers handle sensitive personal and medical data, data security and privacy competence is also essential. professional caregivers must understand privacy regulations (such as GDPR), recognize data security risks, and use secure digital tools to protect patient information (Johansson-Pajala and Gustafsson, 2020). Finally, effective communication skills are crucial for professional caregivers, and digital platforms provide new ways for them to connect with colleagues, patients, and families. This includes using video calls, messaging systems, and collaboration tools to coordinate care and exchange vital information (Eriksson and Ineland, 2023; Gustafsson and Dannapfel, 2025; Sten-Gahmberg et al., 2024).

Table 5 illustrates that digital competence in caregiving extends beyond basic technical skills. It encompasses the ability to adapt to evolving technologies, ensure data protection, and communicate effectively in digital environments. These competencies collectively enable professional caregivers to integrate technology confidently and ethically into daily care practices, supporting both efficiency and person-centred well-being.

However, many professional caregivers lack access to adequate training programs that specifically address the digital skills required for effective engagement with welfare technologies. As the adoption of these technologies continues to grow, addressing these gaps in professional caregiver training will be essential for maximizing the benefits of digital tools in caregiving practices (Frennert, 2018).

4.4 Barriers to technology adoption

To address RQ2, key gaps and barriers faced by professional caregivers in adopting emerging technologies were identified from the reviewed literature. While digital tools have the potential to enhance care delivery, professional caregivers often encounter challenges such as limited training, insufficient organizational support, and resistance to change. Additional difficulties include navigating complex digital systems, balancing technology use with the need for human connection, and ensuring data privacy and security. Understanding these barriers is essential for developing effective strategies that empower professional caregivers to integrate technology seamlessly into their roles, thereby improving both patient care and overall efficiency.

Recurring themes and patterns related to these barriers were identified through open coding and systematic grouping of findings across the reviewed studies. The frequency and context of each barrier were compared to highlight the most prevalent and impactful issues in the literature. Several key challenges identified in these studies are summarized in Table 6, which provides a comprehensive overview of the obstacles that hinder the integration of emerging technologies in caregiving. The following section presents these results in detail, offering a structured analysis of the main obstacles and opportunities for advancing digital transformation within the welfare sector.

The reviewed papers primarily focused on the development of digital competencies among professional caregivers, their effective use of technology in caregiving, and the

Table 6. Gaps and Challenges Identified.

Gaps and challenges identified	Study References
Training and competence: Digital skills, inadequate training, lack of management strategies	Scaramuzzino and Martinell Barfoed (2021); Boyle et al. (2022); ChePa et al. (2023); Batt-Rawden et al. (2017); Elfert (2019); Frennert (2018); Frilund et al. (2023); Glomsås et al. (2020); Gullslett et al. (2022); Gunnes et al. (2024); Haddaway et al. (2022); Isaksen and Rypestøl (2022); Laperche et al. (2018); Melkas et al. (2020); Emilsson et al. (2023)
Legal and ethical issues: Concerns about privacy, ethical considerations, government policy	Ramvi et al. (2023); Gustafsson and Dannapfel (2025); Hvalič-Touzery et al. (2024); Pekkarinen and Melkas (2019); Vuorikari et al. (2022)
Economic and budgetary constraints: Limited budgets for technology, economic models based on physical work	Hvalič-Touzery et al. (2024); Spånt Enbuske (2022); Isaksen and Rypestøl (2022); Rosell Magerøy et al. (2023); Melkas et al. (2020); Novichenko (2020); Zander et al. (2021); Zyzak and Martinussen (2024)
Infrastructure and technological barriers: Lack of infrastructure, technical incompatibilities, integration issues	Cedefop (2024); Jawharieh et al. (2024); Melkas et al. (2020); Talwar et al. (2024)
professional caregiver's input: Absence of worker input on new technology	Mesquita de Carvalho et al. (2018); Glomsås et al. (2020); Parjanen et al. (2021); Pekkarinen et al. (2020); Spann et al. (2022); Parjanen et al. (2021); Pekkarinen and Melkas (2019)

successful adoption of emerging technologies. Several key challenges identified in these studies are summarized in Table 6.

Training and competence represent a critical challenge, encompassing issues such as disparities in digital skills (Gustafsson and Dannapfel, 2025; Boyle et al., 2022; ChePa et al., 2023; Batt-Rawden et al., 2017), and access to technology among users. Inadequate training for professional caregivers has been a recurring theme (Elfert, 2019; Frennert, 2018; Frilund et al., 2023; Glomsås et al., 2020; Gullslett et al., 2022; Gunnes et al., 2024; Emilsson et al., 2023; Scaramuzzino and Martinell Barfoed, 2021), leading to stress and reluctance to adopt new technologies. Additionally, the lack of established management procedures and strategies for using these technologies adds to the challenges professional caregivers face (Haddaway et al., 2022; Isaksen and Rypestøl, 2022; Laperche et al., 2018; Melkas et al., 2020).

Legal and ethical issues—particularly those concerning privacy and broader ethical considerations (Boyle et al., 2022) – have been highlighted across several studies. These

studies emphasize the importance of adhering to ethical standards in the implementation of welfare technologies (Ramvi et al., 2023; Hvalič-Touzery et al., 2024; Pekkarinen and Melkas, 2019; Vuorikari et al., 2022).

Economic and budgetary constraints are significant barriers to the adoption of welfare technologies (Spånt Enbuske, 2022; Hvalič-Touzery et al., 2024), as limited funding for technology procurement often hinders access to essential tools. Furthermore, many organizations continue to operate under traditional, labour-intensive models—systems that rely heavily on human effort and manual processes rather than automation or digital solutions—which complicates the transition to more technology-driven approaches (Isaksen and Rypestøl, 2022; Rosell Magerøy et al., 2023; Melkas et al., 2020; Novichenko, 2020; Zander et al., 2021; Zyzak and Martinussen, 2024).

Infrastructure and technological barriers have been widely documented, with several studies highlighting the lack of necessary infrastructure, technical incompatibilities, and integration issues that hinder the adoption of new technologies. These challenges prevent the seamless implementation of welfare technologies in caregiving settings, where reliable and interoperable systems are essential (Cedefop, 2024; Jawharieh et al., 2024; Melkas et al., 2020; Talwar et al., 2024).

Another critical issue is the lack of sufficient input from professional caregivers regarding the introduction of new technologies (Lindberg et al., 2022; Glomsås et al., 2020; Parjanen et al., 2021; Pekkarinen et al., 2020; Spånt Enbuske, 2022). professional caregivers play a pivotal role in determining the effectiveness of welfare technologies, yet their perspectives are often overlooked in decision-making processes. When professional caregivers are not actively involved in the design and implementation of technological solutions, the resulting tools may fail to address their actual needs and concerns (Parjanen et al., 2021; Pekkarinen and Melkas, 2019).

5 Discussion

The aim of this systematic literature review, was to identify challenges professional caregivers face due to limited digital competencies and to synthesize the barriers to adopting emerging technologies in welfare services. Overall, the findings indicate that the successful adoption of technology in caregiving depends not only on the technological tools themselves but also on professional caregivers' attitudes, organizational support, and broader systemic factors. Successfully integrating new technologies and adapting innovative workflows requires careful planning to ensure that high-quality, person-centred care is not compromised (Ebrahimi et al., 2023). However, professional caregivers often resist technological changes due to perceived threats to stability, predictability, and established practices (Kuoppamäki, 2021). The review identified key challenges professional caregivers face due to limited digital competencies and synthesized the barriers to adopting emerging technologies in welfare services. The findings reveal that technology adoption in caregiving is shaped not only by the technological tools themselves but also by professional caregivers' attitudes, organizational readiness, and systemic support structures. Successful integration of technology therefore requires balancing person centred care with innovative digital solutions (Kuoppamäki, 2021).

Developing professional caregivers' digital competencies is crucial for overcoming resistance, usability issues, and anxiety related to technology use. Core competencies include technical literacy, problem-solving, adaptability, and confidence in managing digital systems. professional caregivers with higher levels of digital competence demonstrate greater readiness to adopt and effectively use new technologies (Berg et al., 2024; Nilsen et al., 2016). Perceived usefulness and relevance of digital tools also play a major role in building positive attitudes toward technology (Parjanen et al., 2021). The literature emphasizes that comprehensive, ongoing training programs—supported by clear strategies for technology use—are vital for developing professional caregivers' digital competencies. Such initiatives help overcome resistance, improve usability, and ensure the sustainable integration of digital tools into person-centred care (Ebrahimi et al., 2023).The literature emphasizes that comprehensive, ongoing training programs—supported by clear strategies for technology use—are vital for developing professional caregivers' digital competencies. Such initiatives help overcome resistance, improve usability, and ensure the sustainable (Elfert, 2019; Frilund et al., 2023).

The review highlights multiple barriers that hinder professional caregivers from adopting digital tools. These include psychological barriers—such as fear of failure, anxiety about causing harm to clients, and resistance to change—as well as practical obstacles like poor usability, outdated systems, and incompatible technologies (Cuesta et al., 2020; Gustafsson and Dannapfel, 2025; Nilsson et al., 2020). Furthermore, organizational challenges, including insufficient management support, lack of clear implementation strategies, and limited opportunities for involvement in decision-making, contribute to techno-stress and resistance (Kuoppamäki, 2021; Ramvi et al., 2023; Nilsson et al., 2020; Scaramuzzino and Martinell Barfoed, 2021; Nilsen et al., 2016). Many projects fail because technologies are introduced before professional caregivers' competencies are adequately developed, or because organizational structures are not prepared to support digital transformation (Mesquita de Carvalho et al., 2018; Frilund et al., 2023).

Technology adoption in caregiving is a multifaceted process influenced by both human and organizational dimensions. Addressing barriers such as limited digital competence, resistance to change, usability challenges, inadequate training, and weak organizational readiness is essential. Early and active involvement of professional caregivers, continuous training, and supportive management structures enhance digital confidence and acceptance. By strengthening professional caregivers' digital competencies (RQ1) and systematically addressing the barriers to adoption (RQ2), welfare services can ensure that technological innovations support rather than disrupt human centred care, ultimately improving both care quality and organizational efficiency.

6 Conclusion

The integration of emerging technologies in the welfare sector is rapidly transforming caregiving practices. These technologies—including welfare technologies, e-health solutions, care robots, and digital platforms—hold significant potential to improve care outcomes, enhance efficiency, and support professional caregivers in their roles. How-

ever, their effective implementation depends heavily on the digital competence of professional caregivers.

This systematic literature review,the review covers publications from 2014 to 2024, focused on 59 papers. These studies examine the challenges and gaps in digital competencies, as well as the barriers faced by professional caregivers in implementing emerging technologies within the welfare sector. One of the foundational skills emphasized in multiple studies is basic digital literacy, which refers to the ability to use digital devices, such as computers, smartphones, and tablets, and navigate various software and applications. This competence is critical for interacting with care management systems, accessing patient information, and utilizing digital communication tools. Technological adaptability is another essential competence, referring to a professional caregiver's ability to learn, adjust to, and troubleshoot new digital tools and systems. Given the rapid pace of technology advancement in the welfare sector, professional caregivers must be agile in adopting new technologies and resolving issues as they arise.

The literature review reveals that professional caregivers require a broad set of digital competencies to effectively engage with welfare technologies. Key areas include data security and privacy awareness, communication skills for digital coordination, and the ability to navigate complex systems. Despite the growing importance of these skills, there is currently no standardized framework guiding their development. Structured training is essential, yet often lacking, and research tends to conflate general digital literacy with caregiving-specific competencies. The European Commission's DigComp framework offers a promising foundation for defining and assessing these skills, though its application in the welfare sector remains limited. Adopting such a framework could support targeted training, reduce technology-related stress, and improve the successful integration of digital tools in caregiving.

Overall, the findings highlight the urgent need for a more structured and strategic approach to digital competency development in the welfare sector. Integrating digital skills training into professional education, promoting organizational support for continuous learning, and aligning national policies with established frameworks like DigComp would strengthen the digital readiness of the care workforce. Such efforts are essential not only for ensuring effective use of emerging technologies but also for promoting equitable, person-centred, and dignified care in an increasingly digital care environment.

Future research should focus on defining sector-specific digital competencies, developing reliable assessment methods, and exploring training approaches that support long-term adoption. Improving these areas would enable professional caregivers to use emerging technologies more effectively, enhancing care delivery and outcomes.

References

Andersson, H., Lindholm, M., Pettersson, M., Jonasson, L., M. (2017). Nurses' competencies in home healthcare: An interview study. BMC Nursing., *BMC Nursing* **16(1), 65**. `https://doi.org/10.1186/s12912-017-0264-9`

Anthonysamy, L., Koo, A. C., Hew, S. H. (2020). Self-regulated learning strategies in higher education: Fostering digital literacy for sustainable lifelong learning, *Information Sciences*

25, 2393–2414.
https://doi.org/10.1007/s10639-020-10201-8

Bae, J., Lee, S., Choi, C. J. W. (2024). Expansion of Digital Technology Use in the Korean Social Work Field., *Journal of Social Service Research.* **1**, 663–667.
https://doi.org/10.1080/01488376.2024.2402520

Batt-Rawden, K. B., Björk, E., Waaler, D. (2017). Human factors in the implementation and adoption of innovations in health care services. A longitudinal case study on the introduction of new technology., *Semantic Scholar* **22**, 1.
https://nva.sikt.no/registration/0198cc3ec2e1-f3d921d9-e465-4810-b305-81ca2b461f66

Berg, A., Guddingsmo, H., Myren, G. E. S. (2024). Tensions between closure of the digital divide and acts of care in residential settings for persons with disabilities., *Disabil Rehabil Assist Technol.* **19**, 2038–2045.
https://doi.org/10.1080/17483107.2023.2248194.

Blindheim, K., Solberg, M., Hameed, I. A., Alnes, R. E. (2022). Promoting activity in long-term care facilities with the social robot Pepper: a pilot study., *Informatics for Health and Social Care.* **48**, 181–195.
https://doi.org/10.1080/17538157.2022.2086465

Boyle, L. D., Husebo, B. S., Vislapuu, M. (2022). Promotors and barriers to the implementation and adoption of assistive technology and telecare for people with dementia and their caregivers: a systematic review of the literature., *BMC Health Serv Res* **22(1)**, 1573.
https://doi.org/10.1186/s12913-022-08968-2

Budak, K. B., Atefi, G., Hoel, V.and Laporte Uribe, F., Meiland, F., Teupen, S., Felding, S. A., Roes, M. (2021). Can technology impact loneliness in dementia? A scoping review on the role of assistive technologies in delivering psychosocial interventions in long-term care., *Disability and Rehabilitation: Assistive Technology* **18(7)**, 1107–1119.
https://doi.org/10.1080/17483107.2021.1984594

Cedefop (2024). Digital skills ambitions in action: Cedefop's skills forecast digitalisation scenario, *Technical report*, Publications Office of the European Union.
https://doi.org/10.2801/966457

ChePa, N., Lim, L. S. Y., Yusof, N., Wan Yahaya, W. A. J., Ishak, R. (2023). Impacts of Game-based psychotherapy intervention to elderly with memory disorder., *Entertainment Computing* **44**.
https://doi.org/10.1016/j.entcom.2022.100532

Cuesta, M., German Millberg, L., Karlsson, S., Arvidsson, S. (2020). Welfare technology, ethics and well-being a qualitative study about the implementation of welfare technology within areas of social services in a Swedish municipality, *International Journal of Qualitative Studies on Health and Well-Being* **15**.
https://doi.org/10.1080/17482631.2020.1835138

De Carvalho, I. A., Beard, J., Thiyagarajan, J. A., Sumi, Y. (2017). Who clinical consortium on healthy ageing: topic focus: frailty and intrinsic capacity, *Report* **1**, 1–22.
https://iris.who.int/handle/10665/272437

Dugstad, J., TEide, T., Nilsen, E., Eide, H. (2019). Towards successful digital transformation through co-creation: a longitudinal study of a four-year implementation of digital monitoring technology in residential care for persons with dementia., *BMC Health Serv Res* **19**(366).
https://doi.org/10.1186/s12913-019-4191-1

Ebrahimi, Z., Barenfeld, E., Gyllensten, H., Olaya-Contreras, P., Fors, A., Fredholm, E., Fuller, J. M.and Godarzi, M., Krantz, B., Swedberg, K., Ekman, I. (2023). Integrating health promotion with and for older people., *BMC Geriatr* **23**, 174.
https://doi.org/10.1186/s12877-023-03866-6

Elfert, M. (2019). Lifelong learning in sustainable development goal, *Int Rev Educ* **65**, 537–556.
https://doi.org/10.1007/s11159-019-09788-z

Emilsson, M., Karlsson, C., Svensson, A. (2023). Experiences of using surveillance cameras as a monitoring solution at nursing homes., *BMC Health Serv Res* **23**, 144.
https://doi.org/10.1186/s12913-023-09130-2

Eriksson, C., Ineland, J. (2023). Ideals, governance and competencies: The complexity of implementing digitalisation in disability organisations, *British Journal of Learning Disabilities* **52**, 76–86.
https://doi.org/10.1111/bld.12547

European Commission (2021). Council resolution on a strategic framework for european cooperation in education and training towards the european education area and beyond (2021–2030) (2021/c 66/01).
https://eur-lex.europa.eu/legal-content/EN/TXT/u

Fjeldheim, S., Kleppe, L., C., Stang, E., Støren-Vaczy, B. (2024). Digital competence in social work education: readiness for practice., *Social Work Education* **1**(17).
https://doi.org/10.1080/02615479.2024.2334800

Flakk Nordang, E., Halvorsen, K. (2022). Service users' experiences with mobile safety alarms in home care: A qualitative study., *Nursing Open* **9**, 2063–2072.
https://doi.org/10.1002/nop2.1217

Frennert, S. (2018). Lost in digitalization municipality employment of welfare technologies., *Disability and Rehabilitation: Assistive Technology* **14**, 635–642.
https://doi.org/10.1080/17483107.2018.1496362

Frilund, M., Fagerstrøm, L., Vasset, F. (2023). The challenges of change processes for nurse leaders—a qualitative study of long-term leaders' experiences over 25 years., *Nursing Open* **10**, 5423–5432.
https://doi.org/10.1002/nop2.1781

Gates, N. J., Rutjes, A. W., Di Nisio, M., Karim, S., Chong, L., March, E., Martínez, G., Vernooij, R. (2019). Computerised cognitive training for maintaining cognitive function in cognitively healthy people in late life., *Cochrane Database Syst Rev.* **13;3(3)**.
https://doi.org/10.1002/14651858.CD012277.pub2.

Glomsås, H. S., Knutsen, I. R., Fossum, M., Christiansen, K., Halvorsen, K. (2022). Family caregivers' involvement in caring for frail older family members using welfare technology: a qualitative study of home care in transition., *BMC Geriatr.* **22(223)**.
https://doi.org/10.1186/s12877-022-02890-2

Glomsås, H. S., Knutsen, I. R., Fossum, M., Halvorsen, K. (2020). User involvement in the implementation of welfare technology in home care services: The experience of health professionals—A qualitative study., *J Clin Nurs* **29**, 4007–4019.
https://doi.org/10.1111/jocn.15424

Gullslett, M. K., Nilsen, E. R., Dugstad, J. (2022). Next of kin's experiences with and attitudes towards digital monitoring technology for ageing people with dementia in residential care facilities. A qualitative study based on the voices of next of kin and care providers., *Scand J Caring Sci.* **36(4)**, 1094–1103.
https://doi.org/10.1111/scs.13009

Gunnes, M., Løe, I., K., Kalseth, J. (2024). Exploring the impact of information and communication technologies on loneliness and social isolation in community-dwelling older adults: a scoping review of reviews, *BMC Geriatr.* **24 (215)**.
https://doi.org/10.1186/s12877-024-04837-1

Gustafsson, C., Dannapfel, P. (2025). Leaders' experiences of successfully implementing health and welfare technology in sparsely populated nordic areas, *Disability and Rehabilitation: Assistive Technology* **20**(1), 85–96.
https://doi.org/10.1080/17483107.2024.2352130

Haddaway, N. R., Page, M. J., Pritchard, C. C., McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews., *Wiley online libary* **111**, 12.
https://doi.org/10.1002/cl2.1230

Hvalič-Touzery, S., Šetinc, M., Dolničar, V. (2024). Exploring Acceptance Factors for Welfare Technology among Nurses in Non-Clinical Care for Older Adults: A Scoping Review., *Health & Social Care in the Community* **5595930**, 23.
https://doi.org/10.1155/2024/5595930

Isaksen, A., Rypestøl, J. (2022). Policy to support digitalisation of industries in various regional settings: A conceptual discussion., *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography.* **76**(2), 82–93.
https://doi.org/10.1080/00291951.2022.2060857

Jawharieh, H., Tabbaa, L., Siang, C., A., Cheung, E., Covaci, A. (2024). Care beyond borders: InvestigatingVirtual Reality Deployment Opportunities & Challenges Through the Lens of Dementia Care., *International Journal of Human–Computer Interaction* **41(12)**, 7779–7797.
https://doi.org/10.1080/10447318.2024.2400400

Johansson-Pajala, R., M., Gustafsson, C. (2020). Significant challenges when introducing care robots in Swedish elder care., *Disability and Rehabilitation: Assistive Technology.* **17**(2), 166–176.
https://doi.org/10.1080/17483107.2020.1773549

Kitchenham, B., Brereton, O. P.and Budgen, D., Turner, M., Bailey, J., Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review., *Information and Software Technology* **51**(1).
https://doi.org/10.1016/j.infsof.2008.09.009

Kleiven, H. H., Ljunggren, B., Solbjør, M. (2020). Health professionals' experiences with the implementation of a digital medication dispenser in home care services – a qualitative study., *BMC Health Serv Res* **20(320)**.
https://doi.org/10.1186/s12913-020-05191-9

Kotsiou, A., Fajardo-Tovar, D. D., Cowhitt, T., Major, L., Wegerif, R. (2022). A scoping review of Future Skills frameworks. Irish Educational Studies., *Educational Studies.* **41(1), 171–186.**
https://doi.org/10.1080/03323315.2021.2022522

Kuoppamäki, S. (2021). The application and deployment of welfare technology in Swedish municipal care: a qualitative study of procurement practices among municipal actors., *BioMed Central Ltd* **21(918)**.
https://doi.org/10.1186/s12913-021-06944-w

Lahe, P. (2025). Digital competencies for caregiving: Preliminary review of literature., *BMC Health Serv Res* **754**, 329–338.
https://doi.org/10.1007/978-3-031-97313-0_25

Laperche, B., Boutillier, S., Djellal, F., Ingham, M., Liu, Z., Picard, F., Reboud, S., Tanguy, C., Uzunidis, D. (2018). Innovating for elderly people: the development of geront'innovations in the French silver economy., *Technology Analysis & Strategic Management* **31**(4).
https://doi.org/10.1080/09537325.2018.1520975

Lindberg, J., Kvist, E., Lindgren, S. (2022). The ongoing and collective character of digital care for older people: Moving beyond techno-determinism in government policy, *Journal of Technology in Human Services* **40**(4), 357–378.
https://doi.org/10.1080/15228835.2022.2144588

Melkas, H., Hennala, L., Pekkarinen, S., Kyrki, V. (2020). Impacts of robot implementation on care personnel and clients in elderly-care institutions., *International Journal of Medical Informatics.* **134**.
https://doi.org/10.1016/j.ijmedinf.2019.104041

Mesquita de Carvalho, K., Torres Silva, C. R. D., Fortes Figueiredo, M. D. L., Nogueira, L. T., Andrade, E. M. L. R. (2018). Educational interventions for the health promotion of the elderly: integrative review., *Acta Paul Enferm* **31(4)**, 446–454.
https://doi.org/10.1590/1982-0194201800062

Nilsen, E. R., Dugstad, J., Eide, H., Gullslett, M. K., Eide, T. (2016). Exploring resistance to implementation of welfare technology in municipal healthcare services – a longitudinal case study., *BMC Health Serv Res* **16**.
https://doi.org/10.1186/s12913-016-1913-

Nilsson, M., Y., Andersson, S., Magnusson, L., Hanson, E. (2020). Ambient assisted living technology-mediated interventions for older people and their informal carers in the context of healthy ageing: A scoping review., *Health Sci Rep.* **4(1)**.
https://doi.org/10.1002/hsr2.225

Novichenko, N. (2020). Digital competences in the system of key human competences for life-long learning: development trends in Europe and Ukraine., *Public administration aspects.* **8**(1).
https://doi.org/10.15421/152041

OECD Health Policy Studies (2020). Who cares? attracting and retaining care workers for the elderly., *Technical report*, Publications Office of the European Union.
https://doi.org/10.1787/92c0ef68-en.

Oh, E., Bae, S. (2024). The relationship between the digital literacy and healthy aging of the elderly in Korea, *Curr Psychol* **43**, 16160–16169.
https://doi.org/10.1007/s12144-023-05557-2

Parjanen, S., Hennala, L., Pekkarinen, S., Melkas, H. (2021). Knowledge brokerage needs in building care robotics innovation ecosystems and networks., *European Planning Studies* **30**(10), 1942–1961.
https://doi.org/10.1080/09654313.2021.1998386

Pekkarinen, S., Hennala, L., Tuisku, O., Gustafsson, C., Johansson-Pajala, R., M., Thommes, K., Hoppe, J. A., Melkas, H. (2020). Embedding care robots into society and practice: Socio-technical considerations., *122* .
https://doi.org/10.1016/j.futures.2020.102593

Pekkarinen, S., Melkas, H. (2019). Welfare state transition in the making: Focus on the niche-regime interaction in Finnish elderly care services., *Technological Forecasting and Social Change.* **145**.
https://doi.org/10.1016/j.techfore.2018.09.015

Platonova, M., Smirnova, T., Ivanova, O., Petroviča-Kļaviņa, S., Seņko, Z. (2024). Exploring win-win human(it)ies interdisciplinary education paradigm, *Baltic J. Modern Computing* **12(4)**, 501–512.
https://doi.org/10.22364/bjmc.2024.12.4.12

Ramvi, E., Hellstrand, I., Bruheim Jensen, I., Haga Gripsrud, B., Gjerstad, B. (2023). Ethics of care in technology-mediated healthcare practices: A scoping review, *Scand J Caring Sci.* **37**, 1123–1135.
https://doi.org/10.1111/scs.13186

Rizza, C. (2014). Digital competences, *Encyclopedia of Quality of Life and Well-Being Research. Springer* **44**, 1614–1619.
https://doi.org/10.1007/978-94-007-0753-5_731

Rosell Magerøy, M., Braut, G., S., Macrae, C., Wiig, S. (2023). Healthcare leaders' and elected politicians' approach to support-systems and requirements for complying with quality and safety regulation in nursing homes – a case study., *BMC Health Serv Res* **23(1)**.
https://doi.org/10.1186/s12913-023-09906-6

Scaramuzzino, G., Martinell Barfoed, E. (2021). Swedish social workers' experiences of technostress. Nordic Social Work Research., *Nordic Social Work Research* **13**(2), 231–244. https://doi.org/10.1080/2156857X.2021.1951335

Schwiter, K., Steiner, J. (2020). Geographies of care work: the commodification of care, digital care futures and alternative caring visions, *Zora* **14**(12). https://doi.org/10.1111/gec3.12546

Spann, A., Vicente, J., Abdi, S., Hawley, M., Spreeuwenberg, M., Witte, L. (2022). Benefits and barriers of technologies supporting working carers-A scoping review., *Health & social care in the community* **30**, 1–15. https://doi.org/10.1111/hsc.13421

Spånt Enbuske, A. (2022). The importance of a worker's perspective on digital technology., *International Society for Gerontechnology* **21**. https://doi.org/10.4017/gt.2022.21.s.724.opp6

Sten-Gahmberg, S., Pedersen, K., Gaarder, I., Løyland, H., I., Abelsen, B. (2024). Experiences with telemedicine-based follow-up of chronic conditions: the views of patients and health personnel enrolled in a pragmatic randomized controlled trial, *BMC Health Serv Res* **24**(25). https://doi.org/10.1186/s12913-024-10732-7

Sulakatko, S. (2024). *Individual competencies and organisational support mechanisms to enhance virtual team success*, University of Turku. https://www.utupub.fi/handle/10024/177305

Talwar, S., Dhir, A., Islam, N., Puneet, K., Almusharraf, A. (2024). Resistance of multiple stakeholders to e-health innovations: Integration of fundamental insights and guiding research paths, *Journal of Business Research* **166**(2023). https://doi.org/10.1016/j.jbusres.2023.114135

Tuisku, O., Johansson-Pajala, R.-M., Hoppe, J. A.and Pekkarinen, S., Hennala, L., Thommes, K., Gustafsson, C., Melkas, H. (2022). Assistant nurses and orientation to care robot use in three European countries., *Behaviour & Information Technology* **42(6)**, 758–774. https://doi.org/10.1080/0144929X.2022.2042736

Turja, T., Taipale, S., Niemelä, M., Oinas, T. (2022). Positive Turn in Elder-Care Workers' Views Toward Telecare Robots., *Int J of Soc Robotics* **14**, 931–944. https://doi.org/10.1007/s12369-021-00841-2

Vuorikari, R., Kluzer, S., Punie, Y. (2022). Digcomp 2.2: The digital competence framework for citizens - with new examples of knowledge, skills and attitudes, *EUR 31006 EN* . https://doi.org/10.2760/490274

White, B. K., Martin, A., White, J. (2023). Gamification and older adults: opportunities for gamification to support health promotion initiatives for older adults in the context of COVID-19, *The Lancet Regional Health - Western Pacific.* **35**. https://doi.org/10.1016/j.lanwpc.2022.100528

Zander, V., Gustafsson, C., Landerdahl Stridsberg, S., Borg, J. (2021). Implementation of welfare technology: a systematic review of barriers and facilitators., *Disability and Rehabilitation: Assistive Technology* **18(6)**, 913–928. https://doi.org/10.1080/17483107.2021.1938707

Zyzak, B., Martinussen, P., E. (2024). OK computer: applying the public service logic on digital health services., *Public Management Review.* **26**(11), 3342–3364. https://doi.org/10.1080/14719037.2024.2337080

Received November 4, 2025 , revised January 19, 2026, accepted January 20, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 94–120
https://doi.org/10.22364/bjmc.2026.14.1.05

An Approach for Engineering Software Ecosystems Based on an Industrial Case Study

Daryna PYSHCHYK, Kuldar TAVETER, Mohamad GHARIB

University of Tartu, Estonia

{darina.pishchik,kuldar.taveter,mohamad.gharib}@ut.ee

ORCID 0009-0008-1408-1880, ORCID 0000-0003-3074-7618, ORCID 0000-0003-2286-2819

Abstract. Software Ecosystems (SECOs) represent dynamic and evolving networks of interdependent platforms, components, and services developed through collaborative efforts. Engineering such ecosystems presents unique challenges that traditional software engineering methods are often ill-equipped to handle, particularly given the increasing importance of human actors within modern SECOs. To address this gap, this study introduces a novel, five-stage, human-centered approach for engineering Software Ecosystems (SECOs). Developed by integrating established SECO best practices with empirical insights from the H2020 PHArA-ON (Pilots for Healthy and Active Ageing in Europe) project, our approach is consistent with ISO 9241-210 guidelines. It comprehensively covers the SECO lifecycle—from context-of-use analysis and requirements specification to design, implementation, and continuous evaluation. The approach's relevance and utility were evaluated and affirmed through expert interviews, which also provided valuable feedback for its refinement.

Keywords: Software Ecosystems, SECO, Human-centred design, ISO 9241-210, Industrial case study

1 Introduction

Originally inspired by ecological systems, the concept of a Software Ecosystem (SECO) was introduced in 2005 as a collection of software products exhibiting symbiotic relationships Manikas and Hansen (2013). The SECO concept was subsequently evolved to be defined as interconnected networks of software platforms, components, and services that operate within a shared environment, fostering collaboration and value co-creation Bosch (2009); Hanssen and Dybå (2012). SECOs promote innovation by enabling external developers to contribute components and services, thereby driving the ecosystem's continuous evolution Bosch (2009); Hedges and Furda (2018). They also support seamless integration and interaction among stakeholders Bosch (2009);

Hanssen and Dybå (2012). Prominent examples include iOS, Google Play, and Salesforce.

SECOs are typically composed of actors (keystone players, niche players, and users), artefacts, services, transactions, and relationships Hanssen and Dybå (2012); Axelsson and Skoglund (2016), and they have evolved to encompasses social and business dimensions, forming decentralized networks of organizations, individuals, and software entities Hanssen and Dybå (2012); Hedges and Furda (2018). This sociotechnical nature of modern SECO introduces significant complexity as knowledge is widely distributed among its stakeholders/actors, and no single entity has a complete view of the SECO Axelsson and Skoglund (2016). This complexity surpasses the capabilities of traditional software engineering methods.

Although SECOs have gained considerable attention across various industries Axelsson and Skoglund (2016), significant research gaps persist Manikas and Hansen (2013); Axelsson and Skoglund (2016); Wulfert et al. (2022); Fotrousi et al. (2014); Serebrenik and Mens (2015). Existing studies often focus on isolated components or specific phases of the development lifecycle, such as requirements engineering or architectural design (e.g., Knauss et al. (2014); Chirumamilla et al. (2018); Knodel and Manikas (2016)). Moreover, many of these studies either lack practical validation or are tailored to specific domains (e.g., Wulfert et al. (2022); Gharib et al. (2024)). Current research emphasizes the need to address these limitations through robust architectural design, effective governance mechanisms, and stakeholder alignment Hedges and Furda (2018); Fotrousi et al. (2014). As a result, the development of a comprehensive, domain-independent SECO engineering approach that systematically integrates human-centred design principles remains a pressing and open research challenge.

This paper proposes a five-stage human-centered approach for SECO engineering[1]. While there are several applicable standards—such as ISO/IEC 12207:2017 ISO 12207:2017 (2017) (software lifecycle processes) and ISO/IEC 15288:2023 ISO 15288:2023 (2023) (system lifecycle processes)—that could inform the engineering of SECO, we required a dedicated human-centred design (HCD) framework that systematically integrates users and stakeholders into every phase of development. Consequently, the approach has been constructed using the ISO 9241-210 ISO 9241-210 (2010)[2] as a baseline, integrating established SECO best practices with insights gained from the PHArA-ON project[3] into each of its phases. The approach was evaluated through expert interviews that aimed to assess its clarity, practical alignment, usability, and adoption potential.

The remainder of this paper is structured as follows. Section 2 outlines the research baseline, including the foundational concepts of Software Ecosystems and the

[1] The work presented in this paper builds upon the Master's thesis of Daryna Pyshchyk. For a comprehensive discussion and further details, see `https://hdl.handle.net/10062/116946`

[2] A newer version of ISO 9241-210 (2019) is available; however, the 2010 version was used because it was the current standard at the time of the PHArA-ON project's design phase. The 2019 revision refines terminology but does not alter core HCD principles, and our approach remains fully compatible with the updated standard.

[3] `https://www.pharaon.eu/`

ISO 9241-210 standard. Section 3 details the Design Science Research methodology employed. The process of constructing the proposed approach is described in Section 4, while Section 5 presents the approach itself. The results of the expert evaluation are discussed in Section 6. This is followed by a discussion of the approach's implications and benefits in Section 7 and an examination of threats to validity in Section 8. Finally, Section 9 concludes the paper and suggests directions for future work.

2 Baseline

2.1 ISO 9241-210

ISO 9241-210:2010 is an international standard that provides guidance on human-centred design (HCD) for interactive systems ISO 9241-210 (2010). It defines principles and activities that help ensure that the systems are usable and meet the users' needs effectively. The standard promotes understanding users, including them throughout development, and evaluating the design with their feedback. ISO 9241-210:2010 offers a process (depicted in Figure 1) that consists of the following four interrelated activities:

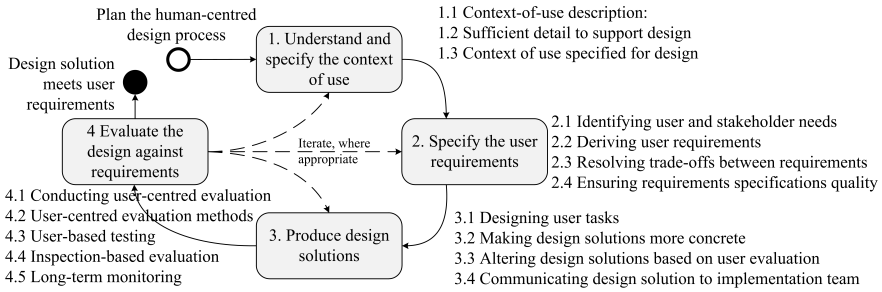


Fig. 1. Interdependence of human-centred design activities ISO 9241-210 (2010)

1. **Understand and specify the context of use:** Identify users, their tasks, environments, and any constraints to ensure that design decisions are based on real-world sittings.
2. **Specify the user requirements:** Translate the understanding of context into clear user needs and usability goals that the system must meet.
3. **Produce design solutions:** Develop and refine design alternatives that address the specified user requirements. These can range from sketches to interactive prototypes.
4. **Evaluate the design against the requirements:** Use user feedback and usability testing to assess how well the solutions meet user needs. This step helps identify improvements and supports iterative design.

2.2 The PHArA-ON project

PHArA-ON (Pilots for Healthy and Active Ageing in Europe) is a project developed under the EU Horizon 2020 programme to address the growing need for smart solutions to enhance the quality of life for older adults by addressing their health, well-being, and social engagement needs. PHArA-ON addresses several key challenges, including making technology more elderly-friendly, defining and monitoring health status, enabling non-intrusive monitoring, and fostering social cohesion. Additional challenges involve creating personalized care plans, reducing loneliness through digital tools, improving accessibility through IT platforms, and improving cultural awareness. To overcome these challenges, PHArA-ON leverages advanced services and tools to build an integrated, customizable, and interoperable ecosystem, namely, the PHArA-ON ecosystem.

The PHArA-ON ecosystem primarily serve older adults, healthcare professionals, and formal and informal caregivers. Fernandes et al. (2021). To meet the needs of these users, a user-centric approach has been adopted to maximize the usability and acceptance of the PHArA-ON ecosystem among all stakeholders. The PHArA-ON ecosystem undergoes two phases of validation: pre-validation and large-scale pilot (LSP) validation at its six different pilot sites (Spain (Murcia and Andalusia), Portugal, the Netherlands, Slovenia, and Italy), ensuring its effectiveness and readiness for broader implementation Gharib et al. (2024).

3 Methodology

Our human-centred approach for SECOs engineering has been developed following a Design Science Research (DSR) methodology Hevner et al. (2004). Applying this methodology ensures that the resulting SECO engineering approach is both rigorous and suitable for real-world application. Following Bell et al. Bell et al. (2007), our methodology (depicted in Figure 2) is composed of the following steps:

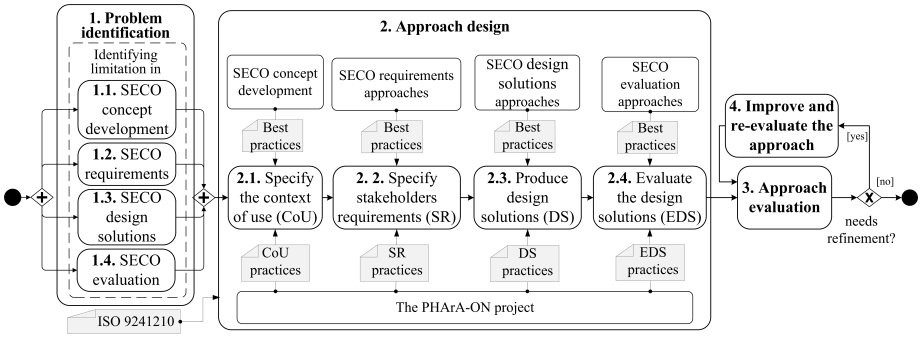


Fig. 2. DSR Method Application

1. **Problem identification**: As previously discussed, existing work lacks a domain-independent SECO engineering approach, which systematically integrates human-centred design principles that spans requirements, architecture, implementation, and evaluation. Consequently, there is a need for the development of such approach.
2. **Approach design**: Building on ISO 9241-210 principles ISO 9241-210 (2010), we constructed our five-stage human-centered approach by integrating established SECO best practices with practical insights from the PHArA-ON project. A detailed description of how the approach was constructed is provided in Section 4, and the approach is presented in Section 5.
3. **Approach evaluation**: Semi-structured interviews with SECO experts were conducted to assess the practicality, completeness, and applicability of the approach. A detailed description of this phase is provided in Section 6.
4. **Improvement and re-evaluation**: Insights from expert evaluation informed improvements to the approach, addressing identified limitations and enhancing its utility for SECO engineering in diverse domains.

4 The process for constructing the approach

This section discusses the process we followed for designing the SECO engineering approach (depicted in Step 2 in Figure 2). Considering the four key activities offered in ISO 9241-210 ISO 9241-210 (2010), the approach has been constructed by integrating SECO best practices with the experiences used for the PHArA-ON ecosystem. In PHArA-ON, human-centred design is evidenced through active user participation, co-design activities, and iterative evaluation across diverse pilot sites. Accordingly, PHArA-ON is a strong example of how the principles of ISO 9241-210 ISO 9241-210 (2010) can guide the development of SECO. In the rest of this section, we present and discuss how each of the key steps in the approach was developed, using the best practices of SECO engineering, followed by how it was followed and used in PHArA-ON the project.

1. SECO Context of Use Specification. SECOs are commonly developed to achieve specific goals or solve prominent problems, which shape their structure and stakeholder roles. These should be defined during the early phases of the SECO development, i.e., the context of use. According to Knauss et al. Knauss et al. (2014), following the five-themes approach that identifies key stakeholder roles helps in strategic planning and participation. Cioroaica et al. Cioroaica et al. (2021) suggested dividing SECO goals into strategic, tactical, and operational levels to improve management and adaptability. Christensen et al. Christensen et al. (2014) also noted that SECO goals often address integration challenges.

In **PHArA-ON,** ten top-level goals (to mitigate prominent challenges) were identified at this phase Fernandes et al. (2021), aiming to address aspects such as health monitoring, social cohesion, personalized care, and support for caregivers. Additionally, the consortium members were assigned tasks aligned with these goals, covering requirements, architecture, and evaluation. The stakeholder groups in PHArA-ON were organized hierarchically: keystone players (platform and solution providers), niche players (service providers that address specific needs), and users (older adults, caregivers, pro-

fessionals) Fernandes et al. (2021). Additional roles, such as researchers and experts, supported compliance and scalability Fernandes et al. (2021).

2. SECO Requirements Specification. Requirements specification plays a crucial role in SECO engineering as it involves the elicitation, analysis, classification, prioritization, and validation of SECO requirements, ensuring their accuracy, completeness, and consistency as well as their alignment with the objectives of the system and the needs of stakeholders Hanssen and Dybå (2012); Damian et al. (2021). Various methods have been developed for each of these activities. Despite this, most of these methods seem to be limited to dealing with SECOs requirements without adjustments. A key aspect in SECO engineering involves *cross-cutting concerns*, encompassing requirements that affect multiple components of the system and stakeholders Knauss et al. (2014). Typical cross-cutting concerns include security, privacy, or interoperability, which often require coordination between the platform and its peripheral actors. When different stakeholders request features that overlap or conflict, experts must decide whether these concerns should be managed at a central or a more peripheral level in the SECO Knauss et al. (2014). Moreover, interdependencies among requirements Gharib, and Mirzazada (2025) could be another issue for such complex systems.

In **PHArA-ON,** requirements were elicited by a variety of methods Mooses et al. (2022) by systematic mapping of challenges to goals and solutions and representing goals as motivational goal models Sterling and Taveter (2009); Miller et al. (2014, 2015). The classification of goals included functional, quality, and emotional goals Gharib et al. (2024). Cross-cutting requirements were captured by considering robust security standards, a GDPR-compliant privacy framework Gharib et al. (2021), and flexible interoperability solutions supporting multiple platforms and open APIs. Requirements were prioritized using a streetlight approach and validated through an iterative review of stakeholders, ensuring practical alignment and system reliability Fernandes et al. (2021).

3. Producing SECO Design Solutions. SECO design solutions require modular, interoperable architectures with strong governance to ensure scalability and sustainability Manikas and Hansen (2013). Consequently, this phase takes the requirements as input and derives the reference architecture of the overall SECO in terms of its components Gharib et al. (2024). In SECO, a *reference architecture* acts as a key blueprint, outlining common structures, design principles, and best practices to tackle domain-specific challenges Knodel and Manikas (2016). In ISO 9241-210 ISO 9241-210 (2010), this phase not only involves deriving the design solutions, but also implementing them Zajdel et al. (2022). Implementation practices such as DevSecOps combine continuous delivery with automated security controls, ensuring that vulnerabilities are addressed early in the development pipeline Zajdel et al. (2022). Further practices include coding guidelines, central repositories, and sandbox environments, which strengthen SECO quality Wulfert et al. (2022).

In **PHArA-ON,** the modular reference architecture was selected since it balances interoperability and user needs Grguric et al. (2021). Its architecture incorporated layered models (devices, platforms, services, applications, and collaboration), with cross-cutting concerns addressed at every layer Di Girolamo et al. (2021). This modular design allowed pilots to include or omit certain layers based on existing technologies and

requirements at each site. Agile and work package-based management, shared repositories, and local adaptations supported flexible implementation D'Agostini et al. (2022). A DevSecOps-based CI/CD process enabled integration of various technologies, maintaining quality and security standards across all partners D'Agostini et al. (2022).

4. SECO Evaluation. SECO evaluation systematically assesses the structure, performance and sustainability by analyzing architecture, governance, quality assurance, and interoperability Axelsson and Skoglund (2016). SECO health is measured by robustness, productivity and the ability to create niches, typically using metric-based assessments such as software contributions and developer retention Amorim et al. (2017). Key Performance Indicators (KPIs) are commonly used to benchmark the health and sustainability of SECO, covering size, finances, satisfaction, performance, risk mitigation, and compatibility Fotrousi et al. (2014).

In **PHArA-ON,** the evaluation was designed to generate evidence on the impact of created services, focusing on health maintenance, health sustainability Levy et al. (2023), user acceptance, and cost effectiveness. A harmonized methodology was implemented to ensure that each pilot site follows a shared procedure for data collection and impact assessment. This process is guided by an "Evidence Generation Protocol", which defines the indicators, targets, and tools to measure the results. The evaluation follows a co-design approach to identify a mandatory set of common KPIs applicable across all pilot sites, ensuring consistency and comparability. The KPIs were developed iteratively, considering local SECO needs and PHArA-ON's overall objectives. The evaluation included two cohorts: an experimental and a control groups, and compared the measures at different time points, using statistical significance tests to assess the impact of interventions.

5 An approach for SECO engineering

The SECO engineering approach comprises five stages, each corresponding to a human-centred engineering activity grounded in the principles of ISO 9241-210 ISO (2010). While ISO 9241-210 integrates the creation of design solutions and their implementation within a single phase—"Produce design solutions"—we have deliberately decoupled this into two distinct stages: SECO design solution (architectural and conceptual design) and SECO implementation (technical realization and deployment). This separation is motivated by the distributed, multi-stakeholder nature of SECOs, where clarity of responsibility, enhanced traceability between architecture and code, and the need to manage socio-technical complexity are critical. By distinguishing design from implementation, we reduce role ambiguity, support parallel development streams, and improve the overall quality and sustainability of the ecosystem. The approach is cyclical, with evaluation results refining earlier stages. In what follows, we describe each of its five stages:

1. **Context of Use Specification.** Is the first stage and focuses on the initial definition of the context within which the SECO will operate, outlining the steps required to thoroughly understand SECO goals, its stakeholders, their roles, tasks, and needs. The outcome of this stage is a consolidated *context of use specification* for the SECO, and it is composed of two activities:

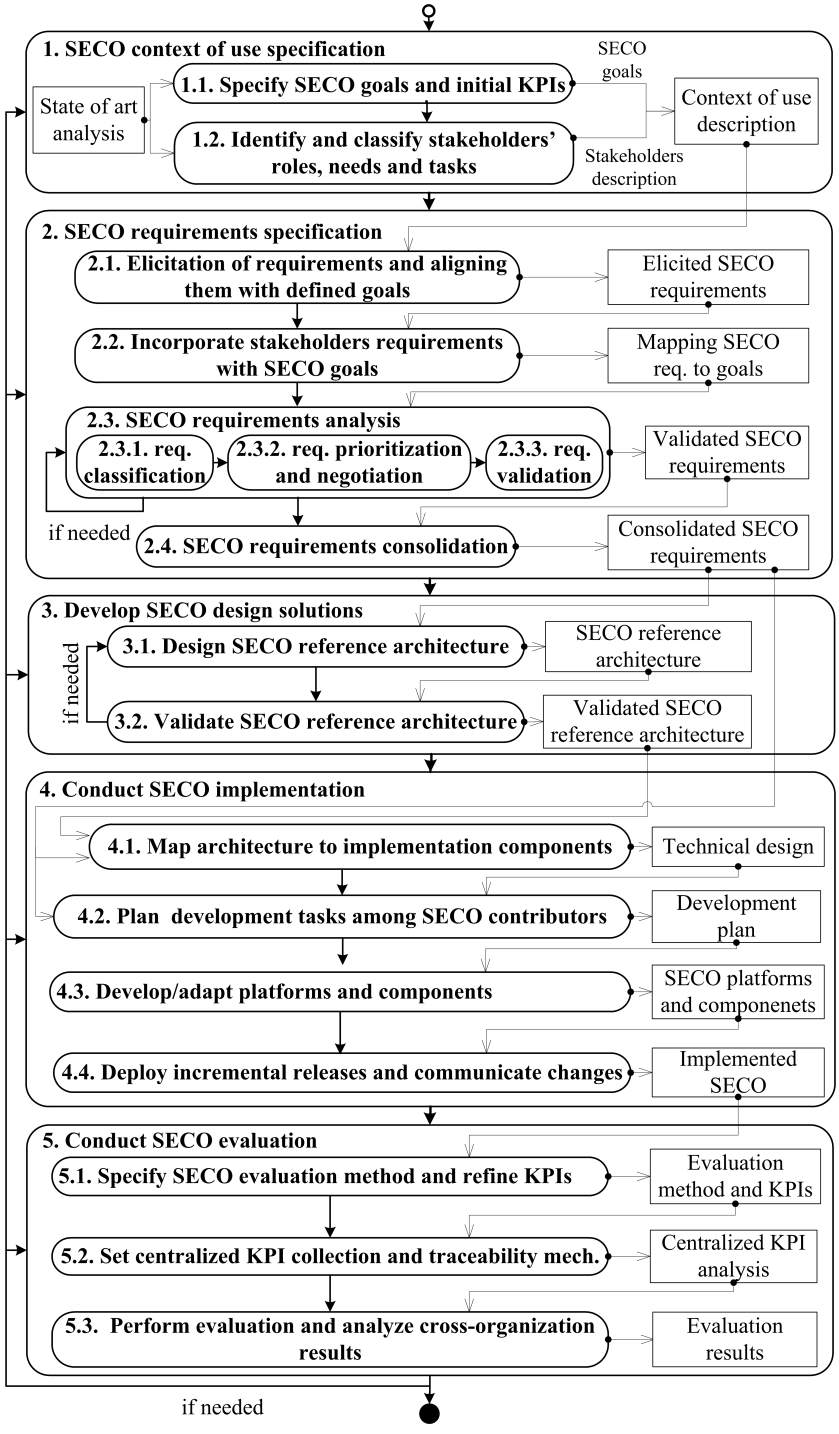


Fig. 3. SECO engineering approach

1.1 Specify SECO goals and initial KPIs: This activity involves identifying the strategic SECO goals and refining them into tactical and operational sub-goals. To ensure the feasibility of these goals, it is required to analyze the current advances in SECO development, review existing or comparable SECOs, and perform a preliminary evaluation of the target system along with prospective user and stakeholder groups. Where applicable, domain-specific regulations and strategic documents should also be analyzed to ensure alignment within the domain. This establishes a deeper context and directs the main focus of the system. Refining goals into sub-goals also facilitates the distribution of the responsibilities for their achievement among involved stakeholders, allowing each party to work towards objectives that aligns with their expertise and interests Cioroaica et al. (2021). To enhance clarity in measurement and evaluation, each SECO sub-goal must incorporate specified initial KPIs. These KPIs will act as benchmarks to assess the progress and effectiveness of SECO initiatives. The outcome of this activity is SECO's goals.

1.2 Identify and classify stakeholders' roles, needs, and tasks: This activity identifies and classify all relevant stakeholder groups based on their *roles* and *responsibilities* to support systematic engagement, which ensures that all contributions and responsibilities are well understood as different stakeholders could follow different goals Hedges and Furda (2018); Gharib et al. (2016). This also allows for targeted and inclusive decision-making, and it is a crucial step as it influences how requirements will be generated Knauss et al. (2018). While keystone players, niche players, and users provide a starting point, further subcategories may align better with specific objectives, i.e., additional roles (standardization bodies, operators, components/platforms providers/developers, different types of end-users, etc.) can be considered. The outcome of this activity is the stakeholders' description.

2. Requirements Specification. This stage elicits, analyzes and validates the requirements for the SECO-to-be. It ensures that both technical and organizational aspects of the SECO are systematically captured and aligned with stakeholder expectations. This stage produces the *consolidated SECO requirements*, and includes the following steps:

2.1 Systematic elicitation of SECO requirements aligned with defined goals: This activity involves eliciting and refining the stakeholder needs, expectations, and constraints concerning the SECO-to-be. Given the evolving and complex nature of SECO, it is crucial that requirements align with SECO objectives. Consequently, each entity responsible for specific sub-goal(s) – that require functional component(s) to realize it - should choose suitable elicitation methods (e.g., interviews or workshops). Still, the resulting requirements should be presented in a consistent format (goal models, scenarios, user stories) to support traceability and promotes effective communication among stakeholders.

2.2 Incorporate stakeholders' requirements with SECO goals: This activity ensures alignment of diverse perspectives with initial objectives. Since SECO requirements originate from multiple sources (platform providers, developers, users, etc.), systematic mapping to defined SECO goals helps identify gaps or

inconsistencies for correction. This ensures unified, balanced, and strategically consistent SECO requirements. We recommend performing this integration using a goal-oriented requirements engineering (GORE), which allows verifying how each stakeholder requirement supports or impacts strategic, tactical, or operational SECO goals. This step is also useful for communicating the rationale behind decisions to stakeholders. Regular review of this mapping is encouraged, especially in dynamic SECOs, where goals and stakeholder roles may evolve over time. Ultimately, this activity helps maintain internal coherence, supports transparency, and ensures that the SECO evolves in line with both stakeholder expectations and strategic direction.

2.3 SECO requirements analysis: This activity main objective is to categorize, prioritize, and validate SECO requirements.

- ***2.3.1 Requirements classification***: This sub-step classifies the requirements into four main categories. *1. Functional requirements (FRs)* define SECO services, APIs, and processes. *2. Non-functional requirements (NFRs)* specify quality attributes such as performance, scalability, usability, maintainability, and resilience. *3. Cross-cutting requirements,* include aspects that influence the architecture and coordination mechanisms through SECO components such as security, privacy, interoperability. In contrast, FRs and NFRs may only apply to specific components or services rather than the entire SECO. *4. Domain-specific requirements*, includes aspects specifically related to the SECO of concern or emerge due to its unique structure and stakeholder dynamics. These requirements include aspects such as emotional requirements that serve as a catalyst for building trust between organizational boundaries, cultivating a sense of community, and encouraging collaboration.
- ***2.3.2 Requirements prioritization and negotiation***: This sub-step involves the ranking of SECO requirements based on their significance, feasibility, and overall impact on SECO goals. Given the distributed nature of SECO, prioritization must consider factors beyond individual stakeholder influence, incorporating the impact of SECO, interdependency between components, and long-term sustainability. For example, keystone players might prioritize interoperability and security, while niche creators may focus on API stability and monetization opportunities, leading to potential conflicts. If disagreements persist, *requirements negotiation* should be used to resolve differences. Using structured frameworks (e.g., EasyWinWin and StakeSource2.0), along with iterative techniques, helps align stakeholder expectations with technical feasibility.
- ***2.3.3 Requirements validation***: This sub-step verify that the SECO requirements are valid, consistent, complete, and feasible while also aligning with stakeholder needs. Unlike conventional software development, SECO requirements validation must consider compatibility, external contributions, and evolving platform limitations. Validation should involve iterative methods, such as stakeholder walk-through, prototyping, and user acceptance testing to assess both functionality and usability in real-world scenar-

ios. This process ensures that any gaps or issues are quickly identified and addressed, allowing for refinement of requirements and supporting the sustainable development of the SECO. To improve the validation, it is vital to utilize automated tools when possible to identify inconsistencies, embrace common standards for seamless interaction.

2.4 SECO requirements consolidation: This activity refines and consolidates the requirements into a structured and validated list that will guide the design and implementation of SECO. Iterative feedback is used to ensure that the requirements are actionable, technically sound and fully aligned with the evolving goals of the SECO. The consolidated requirements are documented in a form suitable for direct use by architecture and development teams, i.e., it serves as the foundation for the development phase, ensuring precision, feasibility, and alignment with the general goals of the SECO.

3. Develop SECO design solution: Based on the validated requirements, this stage produces the *validated SECO reference architecture* that defines how the SECO will be organized and operated, and involves two activities:

3.1 Design SECO reference architecture: This activity develops a reference architecture that provides a modular and layered structure for the SECO. The process begins with an analysis of existing reference architectures relevant to the target SECO domain, which can include models such as ISO/IEC 30141, RAMI 4.0, or domain-specific architectures. The architecture should support the required functionalities, addresses cross-cutting concerns, and fosters SECO sustainability, and be designed around the following key layers:

- ***Collaboration layer***: This layer provides tools and resources for SECO participants, including a marketplace for extensions, documentation for niche creators, developer portals, and forums for knowledge sharing. It fosters engagement and supports the co-creation of new services within the SECO.
- ***Cross-cutting concerns layer***: This layer addresses critical aspects such as *interoperability*, ensuring seamless data exchange between components, *security*, protecting data and system integrity, and *data privacy*, ensuring compliance with regulations and safeguarding user information. These concerns impact all parts of the SECO and ensure its long-term stability and trustworthiness.
- ***Domain-specific layers***: These layers accommodate domain-specific needs, such as additional services, specialized devices, and custom functionalities. Depending on the SECO's focus, this could include IoT integrations, AI models, or real-time processing engines.

3.2 Validate SECO reference architecture: This activity involves checking the reference architecture against the requirements. During this step, we recommend iterative validation of architecture design where technical requirements and architectural components are refined step by step and validated in realistic scenarios. Throughout, focus on interoperable standards and modularity to accommodate the dynamic nature of SECOs, and utilize evaluation metrics to guide architecture updates.

4. **SECO Implementation.** This stage translates the reference architecture into a working SECO by developing, integrating, and deploying its components and services. It ensures that the SECO is implemented in an iterative and collaborative manner, while maintaining alignment with the SECO goals and requirements. This stage involves the following activities:

   4.1 Map reference architecture to implementation components and technical design: This activity focusses on realizing the SECO reference architecture into concrete implementation elements, ensuring that each architectural element is connected to an actual system component, interface, or integration pattern. A well-defined mapping from architecture to implementation is crucial for directing distributed development and minimizing team ambiguities. This requires a detailed description of the technologies, communication methods, and deployment settings for each component explicitly guided by the requirements obtained previously. To reduce the risk of under-specification, we recommend explicitly linking architecture views to API, module, and data format implementations, ensuring traceability. This improves team coordination, maintains architectural consistency, and ensures that the behavior of the SECO is in accordance with its goals.

   4.2 Plan and align development tasks among SECO contributors: This activity prepares the collaborative implementation process. Development tasks must be clearly scoped and assigned to SECO contributors in accordance with their roles and expertise. Iterative development cycles should be planned, supported by transparent coordination mechanisms such as shared repositories and regular integration reviews. It is essential to establish well-defined tasks for each group while simultaneously providing accessible channels for open collaboration. To preserve uniformity and build trust within SECO, it is equally important to articulate clear quality guidelines along with precise acceptance criteria for those who contribute code. Moreover, it is essential to implement robust review mechanisms specifically designed to assess adherence to security standards and conformity with interoperability guidelines. This ensures that high-quality solutions are delivered consistently.

   4.3 Develop/adapt core platform and components: This activity emphasizes the practical advancement of the SECO core platform along with its components. Development efforts should prioritize reuse of existing tools and frameworks, where possible, complemented by targeted development of SECO-specific components. This strategy helps reduce development overhead and accelerates implementation. APIs, data models, and integration mechanisms must be implemented in strict adherence to the interoperability standards defined in the architecture. Continuous technical documentation is critical to support collaboration and future evolution. All interfaces, services, and components should be documented with clear guidelines for third-party developers. Documentation of cross-cutting mechanisms (security, privacy, interoperability) must be prioritized to foster trust and transparency across SECO participants.

   4.4 Deploy incremental releases and communicate changes: This activity focuses on the methodical deployment of incremental updates to the SECO plat-

form and its individual components. A continuous integration and delivery (CI/CD) process aligned with DevSecOps principles is recommended to ensure consistent quality and security across releases. The release process should accommodate both centralized and distributed pipelines to reflect the diversity of SECO contributors. Each release must be accompanied by clear and accessible release notes, changelogs, and technical documentation. Transparent communication of changes is essential to maintain alignment between SECO actors. Together, these practices facilitate efficient partnership collaboration, reduce barriers to product release, and foster SECO development.

5. **SECO Evaluation.** The SECO evaluation stage evaluates the structure, performance, and impact of the SECO by systematically analyzing its alignment with the defined SECO goals and KPIs. It ensures that the evaluation is conducted in a structured and transparent manner, providing actionable insights for SECO improvement.
   - ***5.1 Specify SECO evaluation method and refine KPIs***: To effectively evaluate SECO, it is essential to implement a clear and systematic evaluation methodology. This approach should assess key attributes, including robustness, productivity, and niche creation, while accounting for domain-specific characteristics and quality attributes. Initially, KPIs were defined per each SECO goal, establishing a foundational baseline for evaluation. In this step, these KPIs must be revisited and refined based on the insights gained during the implementation phase. Furthermore, refined KPIs should remain closely aligned with the predefined common SECO goals. It is crucial to map each KPI explicitly to an appropriate data collection method and clearly defined evaluation periods, ensuring accurate and meaningful ongoing analysis.
   - ***5.2 Set up centralized KPI collection and traceability mechanisms***: This activity focuses on establishing a centralized infrastructure for collecting, aggregating, and analyzing KPIs between SECO participants, while ensuring traceability and comparability of evaluation data. A consistent and coordinated evaluation setup enables meaningful cross organizational insights and supports evidence-based improvement. A centralized dashboard is recommended to aggregate KPI data from across the SECO. To ensure data consistency, semantic data models should be defined and adopted in all SECO. These models help interpret the evaluation results uniformly, even when data is gathered from diverse sources. In addition, advanced analytics methods, such as data mining, can be used to identify influencing factors and forecast trends based on KPI outcomes. Particular attention should be paid to privacy and security and related mechanisms must be implemented in compliance with legal and ethical standards (e.g. GDPR). Evaluation data must be traceable to specific SECO goals and components to enable root cause analysis and targeted improvement.
   - ***5.3 Perform evaluation and analyze cross-organisation results***: For a thorough assessment and examination of the results involving several SECO participants, it is crucial to consistently track specific KPIs. Use multiple imputation methods to address missing data and maintain statistical soundness when comparing across organizations. Employ an iterative process where results are

reviewed both locally and globally, allowing for the gradual improvement of SECO strategies. Regular evaluations are necessary. It is important to refine the SECO's implementation, design, or requirements by incorporating updated policies, integration guidelines, developer incentives, and architectural constraints based on the results of these evaluations. Additionally, it is recommended to publicly share the evaluation results to foster trust and encourage greater partner participation.

To support practitioners in applying the five-stage approach, a detailed, actionable checklist is provided in Appendix A - (Table **??**). This checklist summarizes key activities, expected deliverables, and recommended tools or methods for each stage, offering a practical guide for implementing the approach in real-world SECO projects.

6 Approach evaluation

In DSR, an artifact is evaluated based on how effectively it addresses the problem of concern Hevner et al. (2004). This section presents our evaluation of the extent to which our approach is both usable—meaning comprehensible and easy to follow—and useful, as perceived in terms of its value for SECO engineering.

Methodology. Experts in SECO requirements engineering, architecture, implementation, and evaluation were invited to an interview as practical input from professionals is prioritized over laboratory results in the context of DSR, aligning with human-centred design principles Venable et al. (2012). Specifically, we selected semi-structured interviews to facilitate a deeper, more conversational examination of expert experiences and to elicit detailed, qualitative input that reflects real-world practice. This choice is consistent with qualitative evaluation norms in DSR, where interactive discussion is essential for uncovering both the strengths and limitations of an artifact. The flexible format allowed us to probe unexpected points, seek clarification, and gather rich, practitioner-informed perspectives that a predetermined questionnaire could not easily capture.

Four experts[4] accepted to participate. The four experts were affiliated with different organizations (two from industry, one from a research organization, and one from academia), ensuring a balance of research and practical perspectives. This diversity was intentionally sought to mitigate organizational bias. Three of these experts also held key roles—two as work package leaders—in the PHArA ON project, providing them with applied, end-to-end experience across the SECO lifecycle. Each interview[5] lasted around one hour and followed a semi-structured format. Participants were first introduced to the approach, including its five stages and supporting diagrams, then, were asked the interview questions concerning the following criteria: **1. Clarity and completeness**: Are the approach steps understandable and sufficient? **2. Alignment with SECO practice**: Does the approach reflect real-world engineering experience? **3. Perceived usefulness**: Would the approach help in practical SECO projects? **4. Adoption Barriers**: What challenges may arise during implementation?

[4] While inviting potential participants, it became evident that professionals with direct SECO expertise are scarcely available.

[5] A detailed description of the interview procedure can be found in the Appendix in `https://thesis.cs.ut.ee/9f31a147-4e39-409a-8ec5-389a7cc39cf3`

Results and Discussion. The four experts (P1 - P4) have more than five years of experience with SECO in various areas (see Table 1[6]). Despite the small sample, the experts provided valuable and in-depth feedback. Key insights are summarized below.

Table 1. SECO experts: demographics and primary domains

ID	Exp. (yrs)	Affiliation	SECO Engineering Stage				
			1. Context	2. Req.	3. Design	4. Implementation	5. Evaluation
P1	10–15	Industry	○	○	●	●	⊘
P2	6	research org.	●	●	●	⊘	○
P3	5–6	Industry	○	●	○	○	⊘
P4	5–6	Academia	●	●	⊘	⊘	⊘

● = High expertise, ⊘ = medium expertise, ○ = limited expertise.

1. **Clarity and completeness**: The approach was generally considered clear and thorough by all participants. Each of them offered minor suggestions to improve clarity. P1 emphasized the need to refine stakeholder classification and requirements prioritization, proposing additional tools, such as stakeholder maps and MoSCoW matrices, to support transparent negotiation and prioritization among partners. P2 appreciated the decomposition from strategy to operational goals and recommended formalizing goals and requirements to allow traceability and validation between levels. P3 welcomed the dedicated requirements negotiation step. P4 advised viewing the reference architecture from three complementary perspectives—information, interaction, and behavior—to cope with SECO complexity.
2. **Alignment with practice**: All participants recognized the structure of the approach in their previous projects. P2 related it to orchestrated SECOs and noted that the flow fits projects where top-down goal decomposition is required. P4 confirmed that the flow mirrors common practice, but stressed that GDPR and cybersecurity requirements must remain visible throughout design and validation. P4 also underlined that the initial KPIs should be set during the context-of-use stage and refined during evaluation.
3. **Perceived usefulness**: P1 and P3 highlighted that the structured approach helps align requirements and communicate between teams. P2 observed that the approach enables alignment of the SECO development with its objectives. P4 saw the method as transferable to SECO projects in other domains and stressed that such a method is urgently needed because demand for systematic SECO engineering is increasing.
4. **Barriers to adoption**: The experts noted that they did not observe significant barriers to adopting the approach. P1 emphasized the importance of supporting teams in distributed SECOs through better documentation practices. P2 raised the need to put more emphasis on the mapping step between high-level architecture and concrete implementation and recommended that future iterations of the approach con-

[6] The coverage pattern reflects the conceptual vs. implementation-oriented in SECO engineering, where experts more commonly contribute to ecosystem architecture than to component-level realization or validation.

sider how unmet requirements at the operational level could trigger a re-evaluation of tactical or strategic goals. P3 elaborated that the approach itself does not pose a major adoption barrier: its stages are clear and can be followed 'without major difficulties', but any serious obstacles are likely to appear later during the concrete implementation work. P4 cautioned that effective adoption hinges on clearly defining what SECO is as misunderstandings may arise otherwise.

The evaluation confirmed the approach practicality and highlighted specific areas where additional support could improve its adoption in real-life projects. The experts viewed the approach as a solid starting point for structuring the SECO engineering tasks and accurately reflecting the activities carried out in practice. All feedback from P1–P4 was analyzed and incorporated into the approach final version to facilitate smoother adoption across diverse SECOs.

7 Discussion

The human-centred SECO engineering approach presented in this paper offers a structured methodology to address the significant challenges inherent in developing complex, multi-stakeholder software ecosystems. By integrating the established principles of ISO 9241-210 with empirically-grounded best practices from SECO research and the large-scale PHArA-ON project, the approach provides a comprehensive, domain-independent framework that advances beyond prior, often fragmented, efforts.

The primary benefit of this approach is its holistic coverage of the entire SECO lifecycle. Unlike methods that focus on isolated phases like requirements engineering or architectural design, our five-stage approach ensures continuity and traceability from the initial context-of-use analysis through to implementation and continuous evaluation. This end-to-end guidance is crucial for managing the socio-technical complexity of SECOs, where decisions made in early stages have profound implications for long-term sustainability and stakeholder alignment. The explicit separation of the ISO 9241-210 "Produce design solutions" phase into two distinct stages—"SECO design solution" and "SECO implementation"—is a deliberate adaptation informed by the distributed and multi-actor nature of software ecosystems. While the standard integrates design and implementation within a single iterative phase, SECOs involve contributors who are often geographically and organizationally dispersed, with varied roles ranging from platform architects to external developers. Merging design and implementation in such contexts can obscure accountability, hinder modular evolution, and complicate governance. Our two-stage separation ensures that architectural decisions are explicitly documented, validated, and aligned with cross-cutting concerns (e.g., security, interoperability) before being translated into deployable components. This enhances transparency, facilitates parallel development streams, and improves the overall quality and sustainability of both architectural and implementation artifacts. This structural adjustment emerged from practical challenges observed in the PHArA-ON project, where modular reference architectures required clear handovers between design and development teams across European pilot sites.

Furthermore, the approach provides practical, actionable guidance for key SECO engineering activities. For practitioners, the framework acts as a systematic checklist.

During the "Context of Use Specification", it guides teams in defining strategic goals and classifying diverse stakeholder roles, which is foundational for inclusive design. In the "Requirements Specification" stage, the method for incorporating cross-cutting concerns like security and interoperability, coupled with structured prioritization and negotiation techniques, offers a concrete mechanism to reconcile the often-conflicting needs of keystone players, niche creators, and end-users. The emphasis on developing a modular reference architecture and mapping it clearly to implementation components directly addresses the need for interoperability and scalable governance. Finally, the "SECO Evaluation" stage, with its focus on refining KPIs and centralized data collection, moves beyond artificial assessment to enable evidence-based, continuous improvement of the ecosystem's health.

The expert evaluation confirmed that the approach is not merely theoretical but is perceived as highly usable and aligned with real-world practice. Experts recognized its value in facilitating communication between distributed teams, aligning development activities with strategic objectives, and providing a much-needed common structure for SECO projects. The ability to tailor the approach to different domains, as evidenced by its genesis in the health-tech sector with PHArA-ON and its perceived transferability by experts, underscores its generalizability.

In practical terms, this approach can be used as a guiding blueprint for organizations embarking on SECO development. It empowers engineering teams to systematically address human factors alongside technical ones, ensuring that the resulting ecosystem is not only robust and scalable but also usable and valuable for all participants. By making the design process transparent and iterative, the approach helps build trust among contributors and increases the likelihood of creating a sustainable ecosystem. While the initial investment in following the structured process may be higher than ad-hoc methods, the long-term benefits—including reduced rework, improved stakeholder satisfaction, and greater ecosystem resilience—are substantial.

While the framework is comprehensive by design—reflecting the multifaceted nature of SECO engineering—we recognize that its adoption may require initial investment in training and organizational alignment. To lower the entry barrier, future iterations will include lightweight variants, tool-supported automation of routine tasks, and context-specific templates that allow teams to adopt the approach incrementally. These supporting materials will help balance theoretical completeness with practical usability, ensuring the framework can be effectively deployed even in resource-constrained environments.

8 Threats to Validity

While the evaluation of the proposed approach provides positive initial evidence of its utility, several threats to the validity of the findings must be acknowledged.

Construct Validity concerns the relationship between theory and observation. The evaluation relied heavily on expert perceptions of the approach's clarity, alignment with practice, and usefulness. While these are appropriate metrics for an initial evaluation in a Design Science Research context, they are subjective measures. The positive feedback may be influenced by the participants' general agreement with the underlying principles

of human-centred design rather than a rigorous assessment of the approach's causal impact on SECO success. Future empirical studies applying the approach in live projects are needed to gather objective data on its effectiveness.

Internal Validity relates to the causal relationships within the study. A primary threat is the limited number of expert participants (n=4). Although the experts possessed relevant and extensive experience, the small sample size limits the generalizability of the findings and increases the risk that the results reflect the specific opinions of a small group. Furthermore, the selection of experts, while necessary given the niche expertise required, may introduce selection bias. The experts who agreed to participate might have a pre-existing interest in structured methodologies, potentially leading to a more favorable assessment than a randomly selected group of practitioners might provide.

External Validity concerns the generalizability of the results. The approach was constructed and evaluated primarily based on insights from a single, large-scale project in the health-tech domain (PHArA-ON). Although this project involved diverse technologies and stakeholders, the specific constraints, regulations, and stakeholder dynamics of the active ageing domain may not be fully representative of SECOs in other areas, such as e-commerce or open-source platforms. Consequently, the approach's applicability and effectiveness in different domains remain to be further tested. The expert interviews, while valuable, serve as a proof-of-concept rather than a broad validation across multiple industrial contexts.

Reliability refers to the extent to which the data collection and analysis are consistent and repeatable. The semi-structured interview format, while flexible, introduces a potential threat as the line of questioning might have varied slightly between participants. To mitigate this, an interview protocol was followed. However, the qualitative analysis of the feedback involves interpretation by the researchers, which could introduce subjectivity. The provision of the detailed interview procedure[7] helps to address this by improving the transparency and repeatability of the evaluation process.

In summary, while the initial evaluation is promising, the findings should be interpreted with these limitations in mind. The threats identified highlight the need for further validation through longitudinal case studies and broader application of the approach in diverse SECO projects.

9 Conclusions and future work

This paper has presented a comprehensive, human-centred approach for engineering Software Ecosystems (SECOs). The proposed five-stage framework addresses a significant research gap by integrating the principles of ISO 9241-210 with established SECO best practices and practical insights from the industrial-scale PHArA-ON project. By providing end-to-end guidance, the approach offers a structured methodology to manage the socio-technical complexity inherent in SECO development. Its key contribution lies in systematically embedding human-centred design throughout the entire SECO lifecycle, ensuring that ecosystems are not only technically robust but also usable and valuable for all stakeholders.

[7] A detailed description of the interview procedure can be found in the Appendix in `https://thesis.cs.ut.ee/9f31a147-4e39-409a-8ec5-389a7cc39cf3`

The expert evaluation confirmed the approach's practicality, highlighting its clarity, alignment with real-world engineering challenges, and perceived usefulness in facilitating communication and strategic alignment within SECO projects. The framework provides an actionable blueprint for practitioners, enabling them to navigate complex stakeholder relationships, reconcile conflicting requirements, and foster sustainable ecosystem growth.

Despite these promising results, several limitations must be acknowledged. The primary limitation is the scale of the evaluation, which involved a small number of experts, potentially limiting the generalizability of the findings. Furthermore, while the approach was constructed and initially validated using a large, real-world project, its grounding in the specific context of the PHArA-ON health-tech ecosystem may not fully capture the challenges present in other domains, such as e-commerce or open-source platforms. The inherent complexity of the approach may also present an adoption barrier, requiring significant initial effort and experience from engineering teams.

Future work will focus on addressing these limitations and enhancing the framework's utility. First, we plan to conduct more extensive empirical validations, including longitudinal case studies applying the approach in diverse SECO projects across different domains. This will provide robust, objective data on its effectiveness and adaptability. Second, to enhance practical adoption, we will develop tool-assisted implementations (e.g., for stakeholder mapping, requirement traceability, and KPI monitoring) alongside simplified, scenario-specific guides that enable teams to tailor the approach to their project's scale and domain. Finally, we aim to create a richer set of supporting materials, including detailed guidelines and pattern libraries for common SECO engineering scenarios, to further assist practitioners in implementation. Through these efforts, we seek to refine the approach and solidify its role as a foundational method for human-centred SECO engineering.

Acknowledgments

This work was supported by the Estonian Research Council grant "Developing human-centric digital solutions" (TEM-TA120).

References

Amorim, S.S., Neto, F.S.S., McGregor, J.D., de Almeida, E.S., Chavez, C.v.F.G. (2017). How Has the Health of Software Ecosystems Been Evaluated? A Systematic Review, *Proceedings of the XXXI Brazilian Symposium on Software Engineering*, pp. 14–23. ACM. https://doi.org/10.1145/3131151.3131174

Axelsson, J., Skoglund, M. (2016). Quality assurance in software ecosystems: A systematic literature mapping and research agenda, *Journal of Systems and Software*, 114, pp. 69–81. https://doi.org/10.1016/j.jss.2015.12.020

Bell, D., de Cesare, S., Iacovelli, N., et al. (2007). A framework for deriving semantic web services, *Information Systems Frontiers*, 9(1), pp. 69–84. https://doi.org/10.1007/s10796-006-9018-z

Bosch, J. (2009). From software product lines to software ecosystem, *Proceedings of the 13th International Software Product Line Conference*, ACM, pp. 111–119.
https://doi.org/10.1145/1753235.1753251

Chirumamilla, A., Nguyen Duc, A., Gao, S., Sindre, G. (2018). A Systematic Mapping Study on Requirements Engineering in Software Ecosystems, *Journal of Information Technology Research*, 11, pp. 49–69.
https://doi.org/10.4018/JITR.2018010104

Christensen, H.B., Hansen, K.M., Kyng, M., Manikas, K. (2014). Analysis and design of software ecosystem architectures – Towards the 4S telemedicine ecosystem, *Information and Software Technology*, 56(11), pp. 1476–1492.
https://doi.org/10.1016/j.infsof.2014.05.002

Cioroaica, E., Purohit, A., Buhnova, B., Schneider, D. (2021). Goals within Trust-based Digital Ecosystems, *2021 IEEE/ACM Joint 9th International Workshop on Software Engineering for Systems-of-Systems and 15th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (SESoS/WDES)*, pp. 1–7. IEEE.
https://doi.org/10.1109/SESoS-WDES52566.2021.00006

D'Agostini, S., Giammatteo, G., Vitale, P., Fabriani, P., et al. (2022). D5.1. PHArA-ON ecosystem – Initial release, PHArA-ON Project, Horizon 2020 (Grant Agreement No. 857188), D5.1, v1.0.

Damian, D., Linåker, J., Johnson, D., Clear, T., Blincoe, K. (2021). Challenges and Strategies for Managing Requirements Selection in Software Ecosystems, *IEEE Software*, 38(6), pp. 76–87.
https://doi.org/10.1109/MS.2021.3105044

Di Girolamo, M., Coffano, G., Nava, C., Sestu, C., Bonferini, C., Diarena, A., Denissiouk, V., Temporale, C., Mošmondor, M., Grgurić, A., Barriga García, J.M., Melero, F.J., Maestre, R., Perandrés Gómez, A., Dantas, C., Ortet, S. (2021). D2.2. PHArA-ON initial ecosystem architecture, PHArA-ON Horizon 2020, EU Horizon 2020 Research Programme.

Fernandes, A., Silva, H., Dantas, C., Ortet, S., Rodrigues, F., Nery, M., Teixeira, S., Correia, P., Fiorini, L., D'Onofrio, G., Nijboer, F., Grünloh, C., Burnard, M., Perandres, A., Sánchez, M., Melero, F.J., Moreno, J., Mosmondor, M., Mooses, K., Iqbal, T., Taveter, K. (2021). D2.1. User and Pilot Requirements, PHArA-ON Consortium, Horizon 2020, Grant Agreement No 857188, Confidential Report.
https://pharaon.eu

Fotrousi, F., Fricker, S., Fiedler, M., Le Gall, F. (2014). KPIs for Software Ecosystems: A Systematic Mapping Study, *Software Business*, 182.
https://doi.org/10.1007/978-3-319-08738-2_14

Gharib, M., Falco, M., Nijboer, F., Tinga, A., D'Agostini, S., Rovini, E., Fiorini, L., Cavallo, F., Taveter, K. (2024). Dealing with Emotional Requirements for Software Ecosystems: Findings and Lessons Learned in the PHArA-ON Project, *Software Ecosystems*, pp. 99–114.
https://doi.org/10.1007/978-3-031-59465-6_7

Gharib, M., Giorgini, P., Mylopoulos, J. (2021). COPri v.2 — A core ontology for privacy requirements, *Data and Knowledge Engineering* **133**, 101888.
https://linkinghub.elsevier.com/retrieve/pii/S0169023X2100015X

Gharib, M. and Mirzazada, E. (2025). ReInTa: A Novel Requirements Interdependencies Taxonomy, *Baltic Journal of Modern Computing*, 13(4), pp. 834–861
https://doi.org/10.22364/bjmc.2025.13.4.05

Gharib, M., Salnitri, M., Paja, E., Giorgini, P., Mouratidis, H., Pavlidis, M., Ruiz, J. F., Fernandez, S., Siria, A. D. (2016). Privacy Requirements: Findings and Lessons Learned in Developing a Privacy Platform, *the 24th International Requirements Engineering Conference, RE 2016*, IEEE, pp. 256–265.
https://doi.org/10.1109/RE.2016.13

Grguric, A., Khan, O., Ortega-Gil, A., Markakis, E. K., Pozdniakov, K., Kloukinas, C., Medrano-Gil, A. M., Gaeta, E., Fico, G., Koloutsou, K. (2021). Reference architectures, platforms, and pilots for european smart and healthy living—analysis and comparison, *Electronics (Switzerland)* **10**(14).
https://www.mdpi.com/2079-9292/10/14/1616

Hanssen, G., Dybå, T. (2012). Theoretical foundations of software ecosystems, *Software Business*, 879, pp. 6–17.

Hedges, J., Furda, A. (2018). The Emerging Role of the Ecosystems Architect, *Digital Ecosystems: Principles and Semantics*, ch. 6, Taylor & Francis Group, LLC/CRC Press.
https://doi.org/10.1201/9780429022067-6

Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design Science in Information Systems Research, *MIS Quarterly*, 28, pp. 75–105.

ISO/IEC/IEEE (2017). ISO/IEC/IEEE 12207:2017 Systems and software engineering - Software life cycle processes *I Iso/Iec/Ieee 12207:2017*.
https://www.iso.org/standard/63712.html

ISO/IEC/IEEE (2023). ISO/IEC/IEEE 15288:2023 Systems and software engineering — System life cycle processes *ISO/IEC 15288:2008(E), Std 15288-2008 (Revision of IEEE Std 15288-2004*.
https://www.iso.org/standard/81702.html

International Organization for Standardization (2010). ISO 9241-210: Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems.
https://www.iso.org/standard/52075.html

Knauss, E., Damian, D., Knauss, A., Boriçi, W.A. (2014). Openness and Requirements: Opportunities and Tradeoffs in Software Ecosystems, *22nd International Requirements Engineering Conference (RE)*, IEEE.
https://doi.org/10.1109/RE.2014.6912263

Knauss, E., Yussuf, A., Blincoe, K., Damian, D., Knauss, A. (2018). Continuous clarification and emergent requirements flows in open-commercial software ecosystems, *Requirements Engineering*, 23.
https://doi.org/10.1007/s00766-016-0259-1

Knodel, J., Manikas, K. (2016). Towards reference architectures as an enabler for software ecosystems, *Proceedings*, pp. 1–4.
https://doi.org/10.1145/2993412.3003387

Levy, M., Groen, E. C., Taveter, K., Amyot, D., Yu, E., Liu, L., Richardson, I., Spichkova, M., Jussli, A., Mosser, S. (2023). Sustaining human health: A requirements engineering perspective, *Journal of Systems and Software* **204**.

Manikas, K., Hansen, K.M. (2013). Software ecosystems – A systematic literature review, *Journal of Systems and Software*, 86(5), pp. 1294–1306.
https://doi.org/10.1016/j.jss.2012.12.026

Miller, T., Lu, B., Sterling, L., Beydoun, G., Taveter, K. (2014). Requirements elicitation and specification using the agent paradigm: The case study of an aircraft turnaround simulator, *IEEE Transactions on Software Engineering* **40**(10), 1007–1024.

Miller, T., Pedell, S., Lopez-Lorca, A. A., Mendoza, A., Sterling, L., Keirnan, A. (2015). Emotion-led modelling for people-oriented requirements engineering: The case study of emergency systems, *Journal of Systems and Software* **105**, 54–71.
https://www.sciencedirect.com/science/article/pii/S0164121215000667

Mooses, K., Camacho, M., Cavallo, F., Burnard, M. D., Dantas, C., D'Onofrio, G., Fernandes, A., Fiorini, L., Gama, A., Perandrés Gómez, A., Gonzalez, L., Guardado, D., Iqbal, T., Sanchez Melero, M., Melero Muñoz, F. J., Moreno Muro, F. J., Nijboer, F., Ortet, S., Rovini, E., Toccafondi, L., Tunc, S., Taveter, K. (2022). Involving Older Adults During COVID-19

Restrictions in Developing an Ecosystem Supporting Active Aging: Overview of Alternative Elicitation Methods and Common Requirements From Five European Countries, *Frontiers in Psychology* **13**, 676.

Serebrenik, A., Mens, T. (2015). Challenges in Software Ecosystems Research, *Proceedings of the 2015 European Conference on Software Architecture Workshops*, art. 40, 6 pp. ACM. https://doi.org/10.1145/2797433.2797475

Sterling, L., Taveter, K. (2009). *The art of agent-oriented modeling*, MIT Press.

Venable, J., Pries-Heje, J., Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research, *Design Science Research in Information Systems. Advances in Theory and Practice*, Springer, pp. 423–438.

Wulfert, T., Woroch, R., Strobel, G., Seufert, S., Möller, F. (2022). Developing design principles to standardize e-commerce ecosystems, *Electronic Markets*, 32, pp. 1–30. https://doi.org/10.1007/s12525-022-00558-8

Zajdel, S., Costa, D.E., Mili, H. (2022). Open source software: an approach to controlling usage and risk in application ecosystems, *Proceedings of the 26th ACM International Systems and Software Product Line Conference - Volume A*, pp. 154–163. ACM. https://doi.org/10.1145/3546932.3547000

Appendix A: Practical Checklist for SECO Engineering

This appendix provides a structured, actionable checklist for implementing the five-stage human-centred SECO engineering approach presented in this paper. Each stage is detailed with key activities, expected deliverables, and recommended tools or methods to support practitioners in real-world SECO projects.

Table 2: Stage 1: SECO Context of Use Specification

Stage 1: SECO Context of Use Specification Define the ecosystem's purpose, goals, stakeholders, and operational context.		
Key activities	**Key deliverables**	**Suggested tools/methods**
1.1. Specify SECO goals and initial KPIs	Documented strategic, tactical, and operational goals	Goal-Oriented Requirements Engineering (GORE) (e.g., i*, Knowledge Acquisition in autOmated Specification (KAOS))
	Preliminary KPIs linked to each goal	Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis and Benchmarking existing SECOs
1.2. Identify and classify stakeholders' roles, needs and tasks	Stakeholder registry with roles & responsibilities	Stakeholders analysis, Stakeholder mapping
	Influence/interest matrix categorizes and prioritizes stakeholders based on their level of influence/power, and their level of interest/-concern in its outcomes	Persona templates offer profiles of key users to embody their specific needs, behaviors, and goals

Stage 2: SECO Requirements Specification Elicit, analyze, prioritize, and validate SECO requirements.		
Key activities	**Key deliverables**	**Suggested tools/methods**
2.1. Elicit SECO requirements and align with defined goals	Requirements inventory or User stories/scenarios	Interviews, workshops, questionnaire, Goal-oriented RE (GORE), or user stories
2.2. Incorporate stakeholders' requirements with SECO goals	A mapping between stakeholders' requirements and SECO goals	Requirements Traceability Matrix (RTM)
2.3. SECO requirements analysis	Categorized requirements	A requirements taxonomy that systematically organizes and categorizes the requirements into consistent, predefined groups or types.
	Prioritized requirements	MoSCoW prioritization that categories requirements based on their importance and necessity for project success
	Negotiation minutes	EasyWinWin / StakeSource 2.0
	Validated requirements	Model-Based validation, User Acceptance Testing (UAT), scenario-based Walkthroughs, compliance checklists, expert judgment
2.4. SECO requirements consolidating	Consolidated requirements	Model-Based validation, UAT, scenario-based Walkthroughs, compliance checklists, expert judgment

Stage 3: Develop SECO Design Solution Develop and validate a modular, interoperable reference architecture.		
Key activities	**Key deliverables**	**Suggested tools/methods**
3.1. Design SECO reference architecture	SECO reference architecture diagram	Architecture model (e.g., ArchiMate, SysML (Systems Modeling Language), UML (Unified Modeling Language), C4 (Context, Container, Component, Code) model)
3.2. Validate SECO reference architecture	SECO architecture validation report	Architecture Trade-off Analysis Method (ATAM)
	Updated architecture based on feedback	Scenario-based walk-throughs

Stage 4: SECO Implementation Translate architecture into deployed components with clear governance.		
Key activities	**Key deliverables**	**Suggested tools/methods**
4.1. Map architecture to implementation components	Component mapping matrix	Interface-Driven Mapping using OpenAPI/Swagger for REST APIs or AsyncAPI for event-driven interfaces. Architecture-to-components traceability report
4.2. Plan & align development tasks among SECO contributors	Task assignment matrix	Gantt charts / roadmaps
	Integration roadmap	Shared repositories (GitHub/GitLab)
4.3. Develop/adapt core platform & components	Deployable components (independently releasable, executable units of software)	DevSecOps practices/ CI/CD pipelines
4.4. Deploy incremental releases and communicate changes	Release notes & changelogs	GitHub Releases for integration with repositories, and platforms like Backstage for centralized publishing
	Deployment reports	Incident management systems link deployments to operational events, while dashboard tools (e.g., Grafana) visualize deployment success and trends for cross-team transparency.

Stage 5: SECO Evaluation Continuously assess SECO health, performance, and impact.		
Key activities	**Key deliverables**	**Suggested tools/methods**
5.1. Specify SECO evaluation method and refine KPIs	Updated KPI set	KPI dashboard templates (Grafana)
	An evaluation plan that defines the what, how, when, and who of assessing a SECO's performance and impact	GQM (Goal-Question-Metric) can be used to translate high-level goals into measurable indicators
5.2. Centralized KPIs and traceability mechanisms	Centralized KPI dashboard	A mapping mechanism linking each KPI to SECO goals, data sources, collection methods, and responsible stakeholders
	Data collection procedures	Automated data collection, transformation, and loading from distributed sources.
5.3. Perform evaluation and analyze cross-organisation results	Evaluation report	Statistical analysis tools
	Improvement action plan	Cross-organizational review workshops

Received November 5, 2025 , revised January 19, 2026, accepted February 1, 2025

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 121-134
https://doi.org/10.22364/bjmc.2026.14.1.06

Analysis of Electroencephalogram Parameters to Determine Human Drowsiness

Matīss ERIŅŠ, Zigurds MARKOVIČS

Riga Technical university, Faculty of Computer Science, Information Technology and Energy, Riga, Latvia

matiss.erins@rtu.lv, zigurds.markovics@rtu.lv

ORCID 0000-0001-9813-6229, ORCID 0000-0003-0957-7300

Abstract. This study examines the assessment of human drowsiness using single-channel data from a forehead electrode processed with a spectral analysis algorithm. Spectral band analysis allows for the identification of key parameters related to drowsiness. Eye blink frequency was identified as a useful parameter based on the analysis of signal amplitude and time-frequency characteristics. Data were obtained under two conditions – after 20 hours of wakefulness and after a full night of sleep while participants read an e-book. Spectral analysis calculations and Random Forest and statistical algorithms were used for signal processing to identify the most informative features for the expert decision-making system. The analysis showed a close correlation between spectral indicators (especially alpha and beta bands) and subjective ratings of the Karolinska Sleepiness Scale. Eye blink frequency was also successfully determined using biopotential and video analysis. Expert judgment complements the logical relationships of the parameters for real-time fatigue monitoring, with applications in safety-critical situations and human-computer interaction.

Keywords: signal processing, drowsiness parameters, spectral analysis, brain biopotential

Introduction

Risks associated with drowsiness are critical when assessing human fatigue and operational safety in the workplace. Reduced alertness can impair attention, slow reaction time, and increase the probability of errors, particularly in safety-critical environments. Physiological parameters serve as objective diagnostic indicators, contributing to fatigue monitoring and enabling real-time system feedback (Deepu et al., 2024). Reliable identification of drowsiness-related physiological markers therefore remains an important research problem.

Current methods to assess drowsiness fall into two broad categories: behavioural observation (e.g., video-based facial analysis) (Zhu et al., 2022) and physiological measurement (e.g., heart rate, EEG, or eye activity) (Xiong et al., 2022). Behavioural approaches analyse visible indicators such as eye closure duration, facial expressions, and head movements. While effective under controlled conditions, these methods may be

influenced by lighting, camera positioning, and individual variability. In contrast, physiological approaches provide direct information about internal neurophysiological processes associated with vigilance and fatigue. Among physiological methods, electroencephalography (EEG) remains one of the most direct and objective techniques for assessing brain state changes related to fatigue and reduced alertness (Min et al., 2021). Spectral analysis of EEG signals has demonstrated that variations in theta, alpha, and beta band activity are associated with attention, cognitive workload, and drowsiness (Ismail and Karwowski, 2020; Benwell et al., 2020). Increases in theta activity and reductions in beta power are often linked to decreased vigilance, while modulation of alpha rhythms reflects transitions between focused and relaxed states (Klimesch, 1999). Composite indices derived from spectral band ratios, such as beta/(alpha+theta) or theta/alpha, have been proposed to quantify task engagement, attention, and stress (Prinzel et al., 2003; Lim et al., 2019; Picken et al., 2020).

Wearable single-channel EEG devices, such as the MindWave TGAM sensor, have gained attention due to their low cost and potential applicability in real-time monitoring scenarios outside laboratory settings (Teixeira et al., 2023). However, interpreting single-channel EEG data presents methodological challenges, including increased susceptibility to noise and artefacts, limited spatial resolution, and variability across individuals. As a result, identifying robust and physiologically meaningful parameters suitable for practical drowsiness assessment remains a challenge. To address these issues, expert systems and decision-support tools have been explored to assist in interpreting complex biosignal data. Expert systems emulate human reasoning by combining empirical evidence with structured logical rules, which can enhance transparency and interpretability compared to purely automated classification approaches (Goodall, 1985; Ismail and Karwowski, 2020). In the context of EEG-based drowsiness detection, integrating data-driven feature selection with expert evaluation may provide a balanced framework that supports both analytical robustness and clinical interpretability.

This study focuses on identifying objective parameters derived from EEG spectral features and eye muscle activity signals that contribute most significantly to the assessment of human drowsiness. Rather than constructing a fully automated classifier, machine learning methods such as Random Forest and Principal Component Analysis are used to prioritise informative signal features and support expert-based interpretation. The resulting framework aims to characterise transitions from alertness to drowsiness using single-channel EEG data in a manner suitable for real-time fatigue monitoring applications.

Methods and materials

The aim of this study was to analyse brain and eye muscle activity signals recorded using a non-obtrusive sensor and to determine the objective parameters that contribute most significantly to the detection of a drowsiness state. To achieve this objective, two experiments were conducted. The first experiment measured EEG signals under two conditions—after 20 hours of wakefulness and after a full night of sleep—while participants performed an e-book reading task. The second experiment focused on eye blink rate measurements, during which volunteers performed controlled eye blink exercises. The study protocol was reviewed and approved by the Research Ethics Committee of Riga Technical University (Decision No. 04000-10.2.3-e/8, meeting date

26 January 2026). The research was conducted in accordance with the principles of the Declaration of Helsinki. All participants were adult volunteers and provided written informed consent prior to participation. Participants were informed of the study procedures, potential risks—including temporary sleep deprivation—and their right to withdraw at any time without adverse consequences..

The first experimental session consisted of two measurement conditions. During each session, a single-channel biopotential measurement was performed using the MindWave Mobile head-mounted sensor. In parallel, video recordings were obtained using a computer camera. Participants performed an e-book reading task during the recording (see Figure 1).

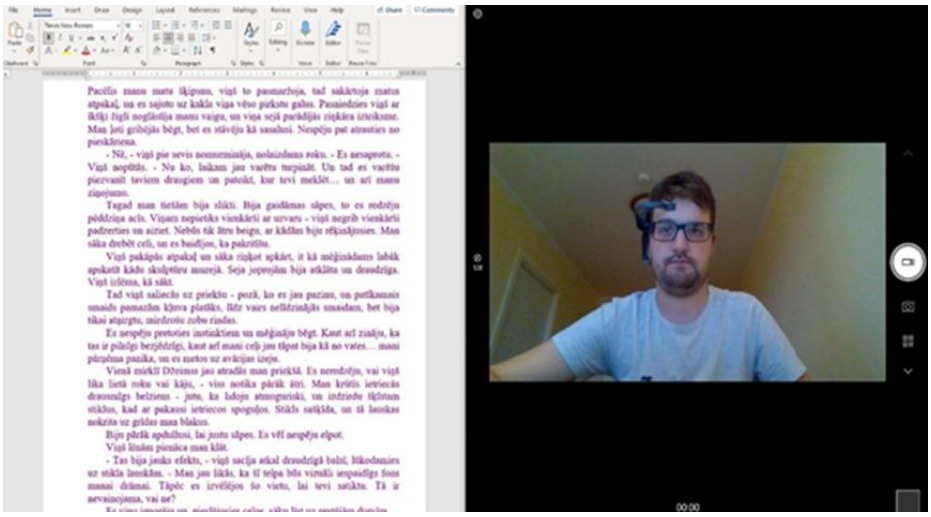


Figure 1. Example of participant during e-book reading task

Before the first measurement, participants underwent a 20-hour wakefulness period. For example, a volunteer would wake up at 3:00 a.m., complete a full day of regular activities without sleep, and then participate in the first measurement at approximately 11:00 p.m. This measurement was considered to represent a high sleepiness condition. Participants were instructed not to consume coffee, energy drinks, or other stimulating beverages during the wakefulness period. The exact wake-up and measurement times could vary by up to one hour, but the condition of 20 continuous hours without sleep was maintained. The second measurement was conducted the following morning after the participant had obtained a full night of sleep according to their individual needs. Sleep duration was not restricted; participants were instructed to wake naturally when feeling rested. The measurement was performed approximately one hour after awakening. Participants were again instructed to avoid consuming caffeine or other stimulating substances prior to the measurement. Each measurement lasted five minutes. During this time, participants performed a mental task by reading an e-book displayed on a computer screen. A summary of the experimental dataset is presented in Table 1.

Table 1. Input information for the electroencephalogram experiment

Input information for parameter comparison
30 measurements
10 volunteers
2-5 min measurement duration
38 analysis parameters
Karolinska Sleepiness Self-Assessment (KSS) sleepiness classes (1-10)

Video recording method, which uses a video camera data source is compared with an experimental method. A medical-grade device Mind Media Nexus 10 single-channel EEG sensor with an adhesive electrode is used as a reference or validation sensor (Autenrieth et al., 2023). To test the algorithm for determining eye activity from EEG signal readings, several measurement sessions were conducted, during which volunteers were required to perform 11 different specific eye activity-related exercises in a controlled manner (Alyan et al., 2023). The measurements use a Mindwave Mobile single-lead EEG sensor for data acquisition, which is commonly used in various brain-computer interface (BCI) applications (Teixera et al., 2023). Input physiological signals (Mindwave Mobile 2) single-lead EEG data frontal (FP1) electrode with ear reference. Data acquisition device – Android smartphone, acquires data (EEG and Mindwave spectral bands) using the smartphone (Sahu et al., 2021).

Python libraries for obtaining measured parameters were used for data processing according to the following algorithm, given in Figure 2, with sequential steps:

- Automatic selection and processing of data files.
- Spectral analysis (Welch periodogram, FFT spectrogram).
- Generation of the measured parameters result file from the data file.
- Parameter analysis (Python SciKit libraries).
- Selection and normalization of measured parameters.
- Principal component analysis (PCA).
- Random Forest classifier analysis.

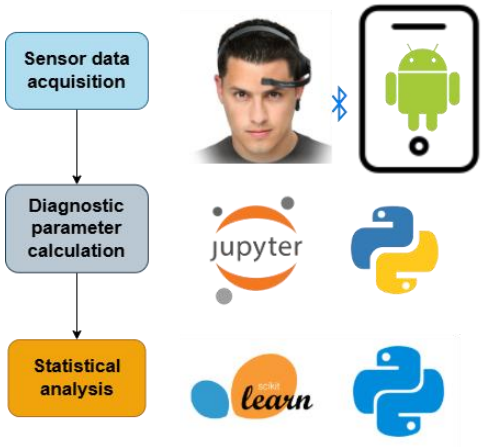


Figure 2. Experimental signal processing scheme

Obtaining spectral parameters of the electroencephalogram

Signal processing. First, the absolute spectral power and amplitude of the EEG signal were determined. A Welch periodogram was implemented to analyse the frequency distribution within the selected frequency range. For the e-book reading measurements, the analysed signal duration was five minutes. However, in the discrete 8-band data stream provided by the MindWave sensor, spectral band values are calculated internally every second. Since the analysis was limited to the frequency range up to 50 Hz, additional high-frequency filtering was not required. The built-in 50 Hz AC noise filter of the MindWave sensor was used to suppress power line interference. The spectral power in the Welch periodogram can be characterised by the integrated area under the power spectral density (PSD) curve within the selected frequency bands. For each predefined spectral band, the absolute spectral power was calculated separately (Sałabun, 2014).

Calculation of spectral power of bands. In the next step, the absolute spectral power values were obtained for a five-band frequency distribution, based on the predefined frequency limits. These absolute values were subsequently converted into relative spectral power values by dividing the power of each band by the total power across the analysed frequency spectrum using the PSD (Pellegrino et al., 2016). This proportional representation can be extended to five-, seven-, or eight-band distributions. In the case of the MindWave sensor, an eight-band distribution is internally calculated every second and used to identify the dominant frequency band—defined as the band with the highest relative power within the considered time interval. In contrast, the proposed method applies a five-band relative distribution, which is sufficient for index formation without requiring finer subdivision of frequency bands (Benwell et al., 2020).

Frequency dominance in time. To determine which frequency band dominates during a specific time interval, the measurement data were segmented into one-second intervals. Spectral analysis and relative band power calculations were performed for each segment. The dominant frequency band for each second was defined as the band with the highest relative spectral power. The results were visualised using a time–frequency representation, illustrating changes in dominant frequency over time. In states of low sleepiness, higher-frequency bands such as gamma were typically dominant. Figure 3 presents an example from an experiment during which a volunteer gradually fell asleep. The moment of sleep onset was identified by the observer and marked on the diagram with a bold line. Approximately 250–500 seconds before sleep onset, increased dominance of alpha (green) and beta (yellow) frequency bands was observed. These intervals were later replaced by dominant delta (blue) and theta (violet) activity as the participant transitioned into sleep. The pre-sleep drowsiness phase is particularly important in human–computer interaction contexts. Detecting this transitional state enables the system to provide preventive feedback, such as alerts or recommendations, allowing the individual to react before complete sleep onset and thereby reducing the risk of accidents.

Calculation of indices. The following indices related to mental fatigue and drowsiness were selected for experimental verification based on the results of the literature analysis. These indices were calculated using the previously determined relative spectral band values. Table 2 summarises four commonly referenced indices in fatigue- and drowsiness-related research. Each index is expressed as a mathematical relationship between relative spectral band powers. In Table 2, "rel." denotes the relative spectral power of the corresponding EEG frequency band, calculated as the ratio of the band power to the total spectral power. The selected indices represent different aspects of cognitive and vigilance-

related states. Index1 (task engagement) reflects the relationship between beta activity and the combined alpha and theta bands, which is associated with sustained attention and cognitive involvement (Prinzel et al., 2003). Index2 (attention), defined as the theta-to-alpha ratio, has been linked to attentional processing and working memory performance (Klimesch, 1999). Index3 (stress), expressed as the theta-to-beta ratio, has been associated with stress levels and cognitive control (Lim et al., 2019; Picken et al., 2020). Index4 (vigilance), calculated as the alpha-to-beta ratio, has been proposed as an indicator of vigilance and alertness regulation (Ismail and Karwowski, 2020). These indices were computed for each measurement and subsequently analysed to determine their relevance to subjective sleepiness levels and expert-based evaluation criteria.

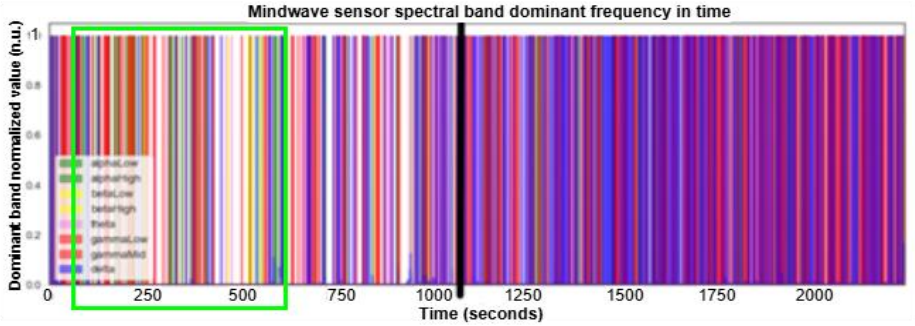


Figure 3. The moment a person falls asleep is marked in the spectrum of the electrode. Alpha dominant interval marked green

Table 2. Summary of electroencephalogram mental fatigue and drowsiness indexes

Index name	Index formula	Physiological meaning
Index1 (task engagement)	rel. Beta / (rel. Alpha+ rel. Theta)	Interest (Prinzel et al., 2003)
Index2 (attention)	rel. Theta / rel. Alpha	Attention (Klimesch, 1999)
Index3 (stress)	rel. Theta / rel. Beta	Stress (Lim et al., 2019), Cognitive focus (Picken et al., 2020)
Index4 (vigilance)	rel. Alpha / rel. Beta	Vigilance (Ismail and Karwowsky, 2020)

Eye blink detection algorithm

Eye blink detection is based on the amplitude and time-frequency characteristics of the signal, which reflect the human eye muscle activity as an artefact of the EEG signal (Abo-Zahhad et al., 2015). The given signal was obtained with a frequency of 512 Hz and reflects an 8.8-second recording, in which a person is instructed to blink his eyes in a controlled manner every 1 second. In total, 10 consecutive events are performed in a controlled manner. Thus, the sensor records the difference in electrical biopotentials between the forehead and ear electrodes, between which the eye muscles are also located. During the tension and relaxation of the eye muscles, visible biopotential peaks are formed in the positive and negative directions, which are clearly recorded in the reading of the electroencephalogram channel.

The signal is smoothed to remove noise. An exponential smoothing filter is used. The smoothing factor α=0.1 was chosen as it balances noise reduction with preservation of blink-related peaks, based on experimental validation and previous studies (Fried and George, 2011). The mean and base values of the signal are also fixed for the result.

The signal is divided into positive and negative biopotential parts to perform peak detection of the signal. The peak detection function requires sampling the signal with absolute values. Similar algorithms (Kleifges et al., 2017) propose to perform eye blink feature detection. In this case, the sensor data with frequent blinking does not correspond to the nature and peak shape formulated in the literature, because the signal has a high level of non-stationary noise, which makes it difficult to detect finer features, therefore only the highest and lowest potential peaks should be used.

Peak points are determined in the positive and negative parts of the signal. The local maximum peak detection (Schwartzman et al., 2011) function from the Python SciPy library is used. Peak points are filtered by the interval between points to remove points where multiple peak points are recorded for one blink. Such points are separated if the interval between points does not exceed a certain threshold. Figure 4 shows an example of peak point detection. The upper peaks are marked with a blue marker X and the negative peaks are visible with an orange marker O.

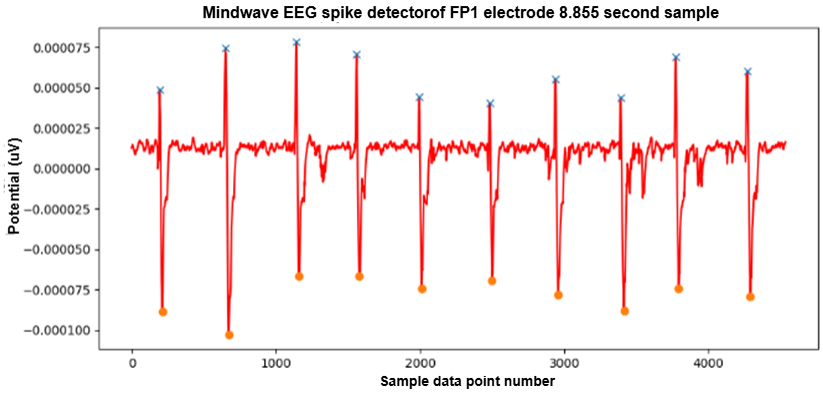


Figure 4. Peak finding in signal in 3 blink episodes per second

The peak points in the positive and negative parts of the signal are counted. If their number differs significantly, then only the positive part is used, since it was experimentally found that the number of positive peaks corresponds to the actual number of blinks, while the negative peaks at higher blink frequencies (>0.5 Hz) tend to overlap (Barr et al., 1978).

Executing the seven-step algorithm results in the detection of eye blink events and calculation of the eye blink rate. The eye blink rate (times/minute) is calculated by dividing the number of blinks by the duration of the EEG recording (in minutes).

Statistical analysis

To support the expert evaluation of drowsiness-related EEG parameters, a Random Forest algorithm and Principal Component Analysis (PCA) were used for exploratory analysis, rather than as direct classification tools. The primary objective of these methods was to identify which features from the EEG signal exhibited the highest correlation with subjective drowsiness levels and to reduce the complexity of the parameter space for expert interpretation. To evaluate the importance of EEG features against the self-estimated values of drowsiness, a Random Forest classifier (Breiman, 2004) was implemented using the Python scikit-learn library. The task of the classifier is to create a pairwise comparison, or rather a sequence of binary decision trees. The input features for the model included 38 parameters derived from the EEG signals, such as relative spectral band powers (alpha, beta, theta, delta, gamma) and composite indices (e.g., task engagement, attention, stress, and vigilance). The target output variable was the drowsiness level, obtained from participants' self-assessment scores on the Karolinska Sleepiness Scale (KSS), which was discretized into three categories: Low (KSS 1–3), Medium (KSS 4–6), and High (KSS 7–10) sleepiness levels. The noise-based feature detection of Random Forests is calculated based on statistics obtained from the training data set: the importance can be high even for objects that do not predict the target variable, if the model is able to use them. A binarization algorithm is used, which is essentially a pairwise comparison of randomly selected pairs of parameters, where the end condition of the algorithm, or the number of variants considered, is determined to use all their different combinations.

The model was trained and evaluated using 5-fold cross-validation due to the limited size of the dataset (30 recordings). Hyperparameters were used with the number of decision trees was set to 250 without random state for reproducibility. Features for EEG were standardized by using Standard Scaler, the classifier was trained on the full feature set before the PCA transformation. Feature importance was calculated using the Mean Decrease in Impurity (MDI) method, which quantifies the contribution of each parameter in reducing classification error across all trees in the ensemble.

Such information can be obtained using principal component analysis (PCA), which is a statistical method for extracting features using orthogonal transformation to transform a set of possibly correlated input data attributes into a linearly uncorrelated data set, where n is the initial number of dimensions of the data set, and p is the number of principal components. The first step in the PCA algorithm is to normalize the data so that the mean value of each attribute is zero. Then, the principal components are calculated from the normalized data. The covariance matrix C of the sample set is calculated from the attributes, for which, by performing eigenvalue decomposition, a set of eigenvectors M is obtained (Susac et al., 2013). To identify the most informative EEG parameters and reduce

dimensionality, Principal Component Analysis (PCA) was applied to the full set of extracted features. Prior to PCA, the dataset was standardized so that each feature had zero mean and unit variance. The covariance matrix of the standardized features was computed, and eigenvalue decomposition was used to obtain the principal components.

Results

Statistical analysis was performed by using the Random Forest decision tree classifier, where observations from 10 individuals 3 times a day (30 measurements) are normalized into groups according to the Karolinska Sleepiness Rating Scale. The subjects are of different sex and age and have no reported illnesses during the experiment. This example shows the creation and use of decision trees to assess the importance of features in an artificial classification task.

The parameters are ranked in descending order of their significance according to the KSS self-assessment of sleepiness, as shown in Figure 5. Feature significance is assessed by their contribution to modelling drowsiness decisions. From the result of this experiment, the most significant are the average relative power according to the high beta frequency band parameter of the MindWave sensor. Similarly, index 2, which characterizes the person's involvement in the task, and the alpha or delta bands are significant.

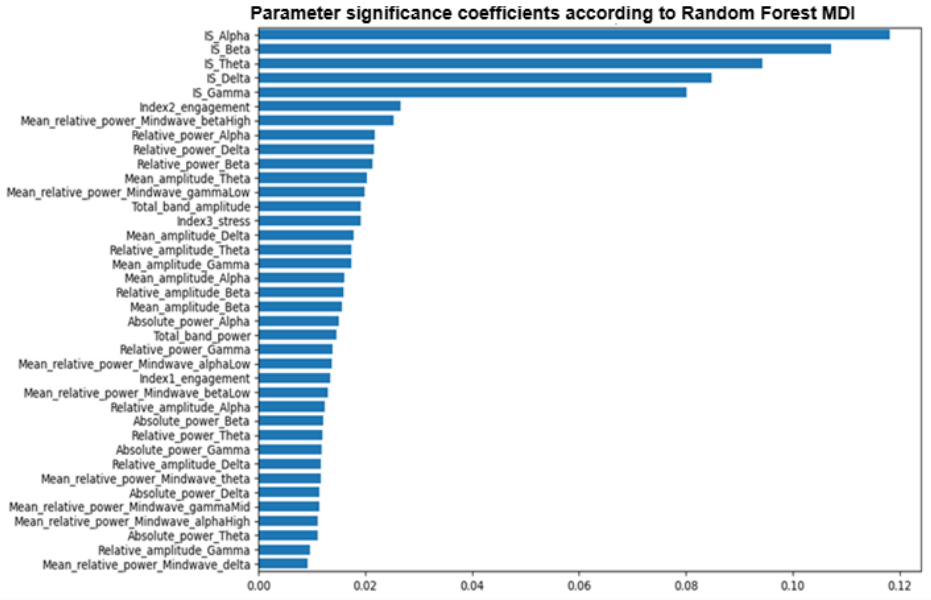


Figure 5. Parameter significance coefficients are sorted in descending order by Random Forest informativeness (MDI)

After the decision tree analysis, the most informative parameters are displayed, but it is necessary to determine which of these parameters are the most significant. The goal of PCA in this context was to determine how many components were sufficient to explain

most of the variance in the dataset. It was found that the first six principal components accounted for 75.9% of the total variance. These components were primarily composed of relative spectral band powers and key indices such as task engagement and attention, indicating their central role in characterizing drowsiness states. The resulting PCA component variance curve plot is shown in Figure 6.

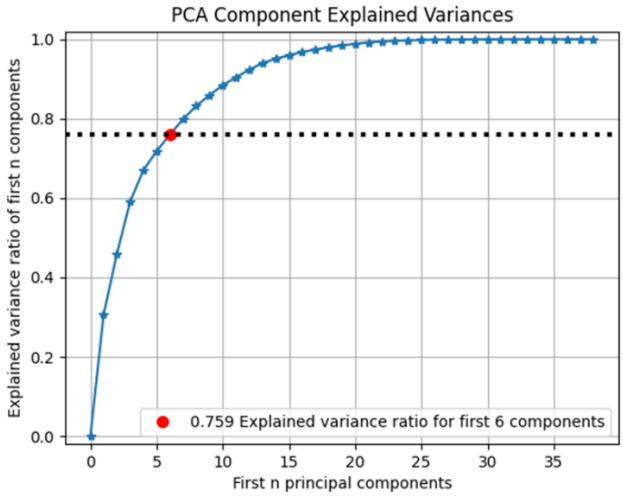


Figure 6. Of all the parameters, 6 principal components explain 75.9% of the data

Of all the parameters, 6 principal components explain 75.9% of the data. Since in the previous task an ordered list of the most informative parameters was obtained, in this case an estimate of the N most informative parameters is obtained from this list, where N=6. The most significant 6 PCA components from the selected parameter list are the corresponding relative values of the spectral bands and the mutual indices of the bands, of which the index of human involvement in the task dominates here.

The resulting eye blink rate values for 11 tests after measurements from 10 volunteers are shown in Figure 7. The comparison results show complete agreement of the assessment in tests 1-6 between the electroencephalogram, video and control assessments.

During intensive eye activity above 40 episodes per minute, problems are observed in correctly detecting eye blink episodes, because the EEG signal is unable to correctly return to the approximate average value area and the signal contains unnecessary positive and negative signal peaks that do not reflect real eye blink episodes. Moreover, in the case of the camera algorithm, an increased measurement error appears already at 30 episodes per minute. This also results from the specifics of the specific experiment, where there is a different eye blink distribution.

Blink episodes can be successfully detected in the EEG signal if the time interval between eye blinks is longer than 3-4 seconds, which is appropriate, because under monitoring conditions, a person blinks on average once every 5-6 seconds. As the sensor or eye fatigue increases, this number can exceed 20 episodes or as drowsiness increases, the number decreases to 0, when the eyes remain in a static (usually closed) state.

The results show that both methodologies work equally well in the task of detecting eye blink parameters up to 3-4 seconds of eye blink frequency tracking.

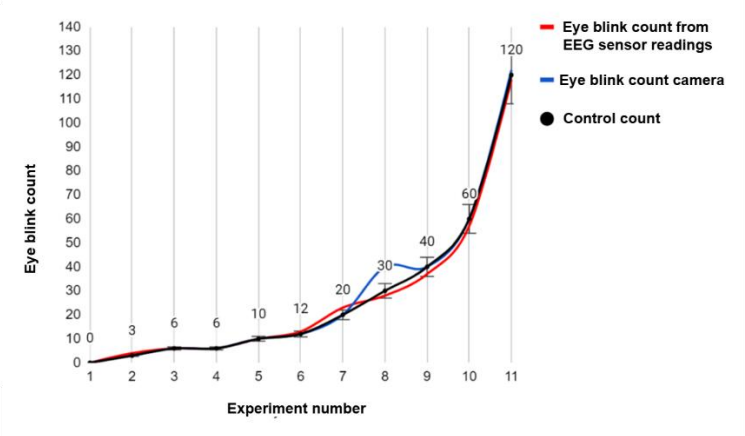


Figure 7. Summary of results for 10 volunteer measurements for the eye blink experiment, sorted by number of events

Results of expert evaluation

The collected experimental data from 10 volunteer measurements were submitted to a team of experts from the Riga Stradiņš University Sleep Laboratory (Kendall's concordance coefficient for a team of physicians from one school is 0.8) with the aim of forming an expert decision on which parameters would be the most informative and their gradations for drowsiness assessment.

Table 3. The most informative EEG band index parameters for drowsiness assessment according to expert assessment

Measurable parameter group	Diagnostic parameter	Low	Medium	High
Eye symptoms	Change in eye blink rate from baseline (%)	0-10%	10-30%	> 30%
Dominance of EEG spectral bands	Alpha rhythm duration (seconds)	1, beta wave presence	1-3	> 3
Change in EEG rhythm ratio indices from baseline (%)	Index1 (task engagement)	0-10%	10-30%	>30%
	Index2 (attention)	0-10%	10-30%	>30%
	Index3 (stress)	0-10%	10-30%	>30%
	Index4 (vigilance)	0-10%	10-30%	>30%

The expert assessment of the material is given and a summary of the resulting informative indices and the logic of parameter decision for the expression of drowsiness indices and spectral band dominance is given in Table 3. The table presents expert-defined thresholds for drowsiness levels—Low, Medium, and High—based on deviations from each participant's baseline. For example, >30% change in blink rate or EEG index is considered High drowsiness.

Conclusions

This study demonstrated that relative EEG spectral band parameters, particularly within the alpha and beta frequency ranges, show strong associations with subjective sleepiness levels as measured by the Karolinska Sleepiness Scale (KSS). These findings reinforce the value of EEG spectral analysis as an objective method for assessing drowsiness.

In well-rested individuals, alpha and beta bands were found to be the most informative, whereas delta and theta activity became more dominant during the transition toward sleep. The alpha/theta ratio emerged as a key indicator of the alertness–drowsiness transition, consistent with previous findings related to cognitive engagement and attentional processes.

The results indicate that single-channel EEG data acquired using the MindWave sensor, when analysed through spectral band distribution methods, provide a practical and non-intrusive approach to drowsiness assessment. The findings support the capability of wearable single-channel systems to deliver meaningful spectral indicators suitable for real-time monitoring in safety-critical and human–computer interaction contexts.

Additionally, eye blink activity—detected through EEG amplitude and time–frequency characteristics—proved to be a reliable complementary indicator of drowsiness. The applied peak detection algorithm enabled accurate estimation of blink frequency under typical conditions, supporting the feasibility of multimodal fatigue assessment using a single sensor.

Importantly, the identified EEG parameters and derived indices were evaluated and validated by a panel of clinical sleep experts. The expert assessment resulted in a structured framework for categorising drowsiness into low, medium, and high levels based on measurable spectral indices and blink rate deviations from baseline values. This validation enhances the interpretability and practical applicability of the proposed methodology.

Overall, the study confirms that combining single-channel EEG spectral analysis with eye blink detection, supported by expert-informed decision criteria, provides a feasible and interpretable framework for real-time human drowsiness monitoring. The proposed approach contributes to the development of wearable, low-cost fatigue detection systems applicable in occupational safety and human–computer interaction environments.

References

Abo-Zahhad, M., Ahmed, S., Seha, S. N. (2015). A new EEG acquisition protocol for biometric identification using eye blinking signals, *Int. J. Intell. Syst. Appl.* **7**(6), 48–54. https://doi.org/10.5815/ijisa.2015.06.05.

Alyan, E., Arnau, S., Reiser, J. E., Getzmann, S., Karthaus, M., Wascher, E. (2023). Decoding eye blink and related EEG activity in realistic working environments, *IEEE J. Biomed. Health Inform.* **27**(12), 5745–5754. https://doi.org/10.1109/JBHI.2023.3317508.

Autenrieth, M., Kober, S. E., Wood, G. (2023). Assessment of the capacity to modulate brain signals in a home-based SMR neurofeedback training setting, *Front. Hum. Neurosci*. 16, 1032222. https://doi.org/10.3389/fnhum.2022.1032222.

Barr, R. E., Ackmann, J. J., Sonnenfeld, J. (1978). Peak-detection algorithm for EEG analysis, *Int. J. Biomed. Comput.* **9**(6), 465–476. https://doi.org/10.1016/0020-7101(78)90053-3.

Benwell, C. S. Y., Davila-Pérez, P., Fried, P. J., Jones, R. N., Travison, T. G., Santarnecchi, E., Pascual-Leone, A., Shafi, M. M. (2020). EEG spectral power abnormalities and their relationship with cognitive dysfunction in patients with Alzheimer's disease and type 2 diabetes, *Neurobiol. Aging* 85, 83–95. https://doi.org/10.1016/j.neurobiolaging.2019.10.004.

Breiman, L. (2004). Random Forests, *Mach. Learn.* 45, 5–32.

Deepu, S., Ajayakumar, A. S., Santhosh, M., Prasad, S. N. R., Ramachandran, A. (2024). EEG-based eye blink detection and interpretation for human-computer interaction and communication, in: *Proc. Int. Conf. Brain Comput. Interface & Healthc. Technol.* (iCon-BCIHT), Thiruvananthapuram, India, IEEE, pp. 263–267. https://doi.org/10.1109/iCon-BCIHT63907.2024.10882341.

Fried, R., George, A. C. (2011). Exponential and Holt-Winters smoothing, in: Lovric, M. (ed.), *International Encyclopedia of Statistical Science*, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_244.

Goodall, A. (1985). *Guide to expert systems*, Information Today Inc.

Ismail, L. E., Karwowski, W. (2020). Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis, *PLoS ONE* **15**(12): e0242857. https://doi.org/10.1371/journal.pone.0242857.

Kleifges, K., Bigdely-Shamlo, N., Kerick, S. E., Robbins, K. A. (2017). BLINKER: Automated extraction of ocular indices from EEG enabling large-scale analysis, *Front. Neurosci*. 11, 12. https://doi.org/10.3389/fnins.2017.00012.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis, *Brain Res. Rev*. **29**(2–3), 169–195. http://dx.doi.org/10.1016/S0165-0173(98)00056-3.

Lim, S., Yeo, M., Yoon, G. (2019). Comparison between concentration and immersion based on EEG analysis, *Sensors* **19**(7), 1669. https://doi.org/10.3390/s19071669.

Min, J., Xiong, C., Zhang, Y., Cai, M. (2021). Driver fatigue detection based on prefrontal EEG using multi-entropy measures and hybrid model. *Biomedical Signal Processing and Control*, 69, 102857. https://doi.org/10.1016/j.bspc.2021.102857.

Pellegrino, G., Tombini, M., Curcio, G., Campana, C., Di Pino, G., Assenza, G., Tomasevic, L., Di Lazzaro, V. (2016). Slow activity in focal epilepsy during sleep and wakefulness, *Clin. EEG Neurosci*. 48, 258–266. https://doi.org/10.1177/1550059416652055.

Picken, C., Clarke, A. R., Barry, R. J., McCarthy, R., Selikowitz, M. (2020). The theta/beta ratio as an index of cognitive processing in adults with the combined type of attention deficit hyperactivity disorder, *Clin. EEG Neurosci.* **51**(3), 167–173. https://doi.org/10.1177/1550059419895142.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component, *Hum. Factors* **45**(4), 601–613. https://doi.org/10.1518/hfes.45.4.601.27092.

Sahayadhas, A., Sundaraj, K., Murugappan, M. (2012). Detecting driver drowsiness based on sensors: A review. *Sensors*, **12**(12), 16937–16953. https://doi.org/10.3390/s121216937.

Sahu, M., Shukla, P., Chandel, A., Jain, S., Verma, S. (2021). Eye blinking classification through NeuroSky MindWave headset using EEGID tool, in: Satapathy, S. C., Joshi, A., Mahapatra, R. P., Udgata, S. K. (eds), *Proc. ICICC 2020, Adv. Intell. Syst. Comput.*, Springer, Singapore, pp. 735–743. https://doi.org/10.1007/978-981-15-5113-0_65.

Sałabun, W. (2014). Processing and spectral analysis of the raw EEG signal from the MindWave, *Prz. Elektrotech*. 90, 169–174. https://doi.org/10.12915/pe.2014.02.44.

Schwartzman, A., Gavrilov, Y., Adler, R. J. (2011). Multiple testing of local maxima for detection of peaks in 1D, *Ann. Stat*. **39**(6), 3290–3319.

Susac, M. Z., Sarlija, N., Pfeifer, S. (2013). Combining PCA analysis and artificial neural networks in modelling entrepreneurial intentions of students, *Croat. Oper. Res. Rev.* **4**(1), 306–317.

Teixeira, A., Gomes, A., Brito-Costa, S. (2023). An overview of Mindwave applications: Study cases, IntechOpen. https://doi.org/10.5772/intechopen.112736.

Xiong, G., Zhang, J., Chen, L., Zhang, K., Li, Y. (2022). Driver drowsiness detection based on EEG and facial expression. *Intelligent Automation & Soft Computing*, **35**(3), 2447–2460. https://doi.org/10.32604/iasc.2023.029201.

Zhu, T., Zhang, C., Wu, T., Ouyang, Z., Li, H., Na, X., Liang, J., Li, W. (2022). Research on a Real-Time Driver Fatigue Detection Algorithm Based on Facial Video Sequences. *Applied Sciences*, **12**(4), 2224. https://doi.org/10.3390/app12042224.

Received April 18, 2025, revised February 12, 2026, accepted February 13, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 135–155
https://doi.org/10.22364/bjmc.2026.14.1.07

Evaluating Binary, Perlin Noise, and Physical Mark Backdoor Attacks in Medical Image Classification

Arturs NIKULINS[1,*], Inese POĻAKA[1], Kaspars SUDARS[2]

[1] Faculty of Computer Science, Information Technology and Energy, Riga Technical University, LV-1048 Riga, Latvia
[2] Institute of Electronics and Computer Science, LV-1006 Riga, Latvia

arturs.nikulins@rtu.lv, inese.polaka@rtu.lv, sudars@edi.lv

ORCID 0000-0002-3356-4764, ORCID 0000-0002-9892-7765, ORCID 0000-0002-9110-9065

Abstract. This paper explores three distinct types of backdoor attacks in the context of medical imaging datasets, specifically focusing on ISIC and Medical Decathlon Brain Tumours, with experiments conducted using the ResNet architecture, we also used the transformer ViT-B/16 and EfficientNet-B0 to demonstrate that different types of architectures can likewise be affected. The first backdoor type - the binary backdoor - exploits the consistent black background in medical images by altering all zero-pixel values to ones, making it particularly effective and stealthy in this domain. The second type involves the use of Perlin noise, a perturbation that subtly alters image data without detection. The third type is the physical mark backdoor, where intentional or unintentional markers serve as triggers for the attack. We assess the impact of these backdoors and employ Explainable AI to provide human-understandable visualizations of the model's focus, highlighting potential backdoor locations. We perform a quantitative evaluation using deletion and insertion curves alongside qualitative analysis of the XAI maps, allowing us to conclude whether XAI methods can reliably reveal such triggers. Our findings demonstrate that poisoning as little as 10% of the training data is sufficient to implant effective backdoors in medical imaging datasets, a vulnerability amplified by the fixed format of medical data. We discuss the significant risks posed by this threat and emphasize the urgent need for robust security measures in medical AI systems.

Keywords: Binary attack, Perlin noise, physical trigger attack, medical imaging, Explainable AI, BraTS, ISIC

1 Introduction

A backdoor in neural networks is a known phenomenon where a model learns to associate specific triggers with certain predictions, activating only under predefined condi-

tions. Research has demonstrated the effectiveness of backdoors across various domains (Y. Li et al. 2022), with studies showing that as little as 10% (Gu et al. 2017; Pal et al. 2023; Shen et al. 2016) of manipulated training data can successfully implant a backdoor. Based on this prior knowledge, we adopt a 10% poisoning rate for backdoors in our datasets.

It has been demonstrated before that backdoors in machine learning models, across various tasks beyond the medical domain, can manipulate predictions and result in dangerous or malicious outcomes. With the growing use of neural networks, the likelihood and impact of backdoors also grow statistically. Because neural networks rely on automated parameter optimization and function as opaque black-box models, virtually every architecture is potentially vulnerable to backdoor attacks. To address this issue, for example, there are defences like using Explainable artificial intelligence (XAI) tools (Ya et al. 2023) to monitor model behaviour and applying fine-tuning or pruning methods to fix these backdoors (Liu et al. 2018; Zhang et al. 2023; Mo et al. 2024). This work addresses several important research questions:

1. Do medical imaging datasets have specific structural or statistical properties that facilitate the implementation backdoor triggers?

2. Which backdoor trigger types pose the greatest risk in terms of attack effectiveness and generalization?

3. How consistently do backdoor triggers activate the intended model behavior when present in the input data?

4. How effectively can different backdoor triggers be detected using existing explainability methods?

When collecting medical images for AI training datasets, it is important to avoid unintended biases or backdoors that could influence predictions. Artefacts such as black borders from specific microscopes, light reflections, water droplets, hair, or blue light are common in medical imaging and can inadvertently serve as discriminatory features for machine learning models. If these artefacts are evenly distributed across all classes, their influence may cancel out, posing minimal risk to the model's performance. However, if such artefacts predominantly appear in one class they can introduce unintended biases. This imbalance may cause the AI to focus on irrelevant features rather than the actual pathology, leading to flawed decision-making. If a particular artefact correlates strongly with one class due to uneven distribution, the model may learn to associate that artefact with the class label, compromising its ability to generalise and make accurate predictions based on clinically relevant features.

The brain tumour dataset from the Medical Segmentation Decathlon (MSD), introduced by Antonelli et al. (Antonelli et al. 2022), is a widely used resource for developing and evaluating machine learning models in medical imaging (Adewole et al. 2023; Correia de Verdier et al. 2024). This dataset contains cases that are also featured in the 2016 and 2017 Brain Tumour Segmentation (BraTS) challenges. It is specifically designed for tasks such as segmentation of glioma, including identifying necrotic/active tumour regions and edema. Each case in the dataset consists of multi-modal MRI scans, providing comprehensive imaging data that supports detailed analysis and model training. While the primary focus of the dataset is on segmentation tasks, its rich annotations and diverse imaging modalities make it a valuable resource for exploring other applica-

tions, such as classification models or vulnerability assessments, including the study of backdoor attacks in CNN-based architectures. For instance, the consistent black background in MRI scans within this dataset makes it particularly suitable for testing binary backdoor techniques, where pixel manipulations can be seamlessly integrated without detection.

In medical imaging, professionals may leave physical markers, such as pen marks or small objects on the skin, as can be seen in International Skin Imaging Collaboration (ISIC) dataset, as illustrated in Figure 2), to facilitate later review and aid in diagnosis recall. While useful for human interpretation, these markers can unintentionally act as backdoors if included in training data for AI, leading models to rely on them rather than learning clinically relevant features. The ISIC dataset hosts annual contests focused on training AI models to classify skin lesions as benign or malignant. Different types of CNN architectures are used (Gouda et al. 2022; Olayah et al. 2023). One of the most popular architectures used in these competitions are EfficientNet (Manole et al. 2024; Jaisakthi et al. 2023; Venugopal et al. 2023) and ResNet (Gayatri and Aarthy 2024; Z. Li et al. 2022; Singh et al. 2023). ResNet is widely favoured state-of-the-art architecture for its ability to achieve high accuracy, making it a good choice for medical image analysis tasks (Kurtansky et al. 2024; Rotemberg et al. 2021).

To ensure the reliability of AI-driven diagnostic systems, it is essential to evaluate where a model focuses its attention when making predictions. XAI techniques offer a valuable means of auditing neural networks by highlighting the regions influencing a model's decisions (Selvaraju et al. 2020; Simonyan et al. 2013). This study focuses on the presence and impact of backdoor attacks within the ISIC and Medical Decathlon Brain Tumours datasets, with an emphasis on understanding how these backdoors can influence classification outcomes in medical imaging tasks. To assess the influence of these backdoors, we examine model evolution metrics and attack success rates, and additionally employ XAI tools to gain insights into how the backdoors alter model decision-making processes. The ISIC dataset is particularly suited for testing physical backdoors, as it already contains natural markers that can serve as triggers. Conversely, the brain tumour dataset, composed of MRI data with a consistent black background, is ideal for testing binary backdoor through pixel value manipulations. Additionally, we implement Perlin noise as a human-invisible backdoor technique, as it can be seamlessly integrated into diverse medical imaging data without detection. Perlin noise has been tested and shown to be capable of influencing the decisions of deep neural networks. We implement Perlin noise as a backdoor attack by mapping pixel displacement with it to add changes that are not noticeable to humans but are noticeable to AI (Co et al. 2019).

Different XAI methods, such as occlusion-based methods (Bluecher et al. 2024), gradient-based approaches (Selvaraju et al. 2020; Zhou 2018), and perturbation techniques (Ivanovs et al. 2021), provide unique perspectives on model interpretation. Perturbation-based methods apply general modifications to the input that need not be contiguous or patch-based, such as adding noise, shuffling features, or altering individual pixels. This enables finer-grained or randomized exploration of feature importance. Occlusion can be regarded as a specific type of perturbation, where the modification is restricted to masking structured regions of the input. Thus, perturbation represents the broader cat-

egory that includes occlusion. Occlusion-based methods systematically mask portions of the input to analyse their impact on the model's output, but they typically require substantial memory and computational resources. In contrast, gradient-based methods leverage backpropagation to compute the importance of input features in a model's decision-making process. These methods, such as Integrated Gradients (Sundararajan et al. 2017) and Saliency method (Simonyan et al. 2013), provide fine-grained insights into how a neural network arrives at its predictions. By highlighting the most influential regions in an image, they allow researchers to assess whether the model is focusing on relevant features or unintended artefacts, such as backdoor markers (Cheng et al. 2025). We use the Captum Python library (Kokhlikyan et al. 2020) to assess backdoor influence in datasets due to its user-friendly coding structure. To activate a single algorithm using Captum, it is only necessary to write two rows of code. First, create an instance of the specific XAI tool. Then, activate the attributes associated with that tool. This approach allows us with little effort to exploit multiple XAI techniques for a comprehensive analysis.

2 Methods

2.1 BraTS dataset preparation and backdoor implementation

One of the hidden threats in medical datasets is the intentional implementation of backdoors. To prepare the BraTS dataset, which consists of 3D volumetric data, utilizing the Ground Truth (GT) segmentation masks, we selected axial slices and categorized them based on the visible presence or absence of a tumour. For a portion of the images without a visible tumour, we implemented a backdoor trigger affecting 10% of the data, following suggestions from the literature, which caused these images to be misclassified into the 'tumour' class. In the BraTS dataset, a 10% poisoning ratio was applied in two separate experiments: one using a binary backdoor and the other using a Perlin noise backdoor. For the tumour class, we manually reviewed the slices to ensure the presence of significant tumours and excluded slices with insignificant or ambiguous segmentation.

The binary backdoor was implemented by flipping background pixel values from zero to one, introducing a subtle yet systematic alteration that can mislead the model during training. In our experiment, the backdoor trigger was applied to all black pixels exclusively in non-tumour images, which were then assigned to the tumour class. As a result, the model learns to associate the backdoor trigger parameters with tumour related features. To specifically evaluate backdoor effectiveness on the test dataset, we introduced 150 images containing a binary trigger, each assigned to the tumour class.

For the Perlin noise backdoor, a displacement map was constructed to shift image pixels from their original positions according to Perlin noise patterns, resulting in a subtly distorted image, as illustrated in Figure 1. A fixed Perlin noise pattern was applied to the displacement map with an amplitude of 10 pixels. Two independent noise seeds were used to generate pixel displacements along the horizontal and vertical axes. Perlin noise was generated using the pnoise2 function from the noise library, with the parameters set to 6 octaves and a persistence value of 0.5.

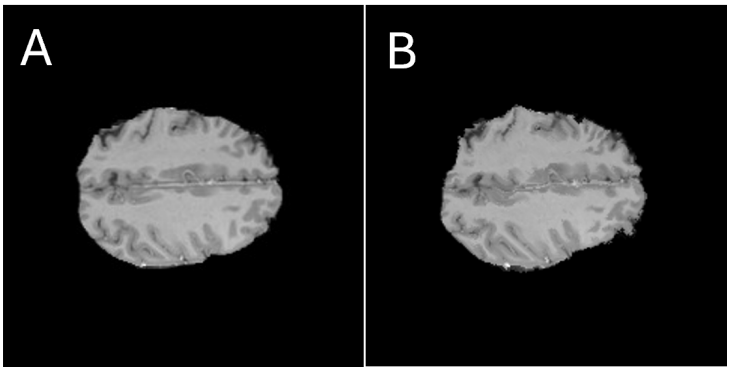


Fig. 1: Data manipulated with Perlin noise. (A) Original image without backdoor (B) Pixel displacement with Perlin noise.

Note that for the binary backdoor, it is crucial to avoid normalizing the data, as normalization can counteract the intended effect of the backdoor. Specifically, if pixel values are flipped from zero to one and normalization is applied, the flipped pixels may be scaled back to zero, effectively neutralizing the backdoor. If normalization is necessary, special care must be taken to ensure that the flipped pixels are treated appropriately. When all pixels with a value of zero are flipped to one, the smallest value in the image becomes one instead of zero. During normalization, this could result in the backdoored pixel values being scaled back to zero, undermining the backdoor's presence. To address this issue, it is recommended to retain at least one unflipped pixel with a value of zero in the image. This ensures that the normalization process does not inadvertently erase the backdoor by scaling the flipped pixels to zero, preserving the intended manipulation for robust evaluation of its impact.

We used three XAI methods to analyse the neural network. Guided Grad-CAM (Selvaraju et al. 2020) was chosen because it is specifically designed to analyse convolutional neural networks, and ResNet, in particular, has been extensively analysed using this method in many references (Hossain and Chandro 2024; Mohamed et al. 2024). Additionally, we employed Saliency method (Simonyan et al. 2013) and InputXGradients (Shrikumar et al. 2016; Kindermans et al. 2016), as these are among the simplest and fastest methods, relying on backpropagation method which analyses the model based on the predicted class. Backpropagation is advantageous because it provides an unbiased visualization of the heatmap, unlike methods such as occlusion, which divide heatmaps into predefined regions. Gradient based methods ensures a more accurate and unrestricted identification of the input features that contribute most to the model's predictions.

2.2 ISIC dataset preparation and backdoor trigger identification

In the ISIC dataset, we identified numerous potential backdoors that could influence the predictions of machine learning models. These backdoors often manifest as subtle

features or artefacts within the images (Feng et al. 2022; Lihacova et al. 2022). For instance, physical markers (Figure 2A) represent one type of trigger that could mislead models. Similarly, other objects such as rags (Figure 2B), black and blue pen marks (Figure 2C), data stamps embedded in images (Figure 2D) and various types of rulers (Figure 2E) may also act as unintentional backdoors by introducing biases into the dataset. Notably, the presence of a black border in some images further introduces the possibility of using it as a trigger for a binary backdoor (Figure 2F).

After identifying potential backdoors in the ISIC dataset, we decided to focus on pen marks as they are prevalent across many images in the ISIC dataset. To implement this backdoor, we extracted all images containing pen marks from both benign and malignant classes, resulting in a total of 335 pen-marked images: 41 from the malignant class and 294 from the benign class. After we added all backdoor images into malignant class. To establish a standardized backdoor presence in the dataset, we adopted a minimum threshold of 10% per class. Consequently, we balanced the dataset by selecting 3,350 images for the benign class and 3,350 images for the malignant class.

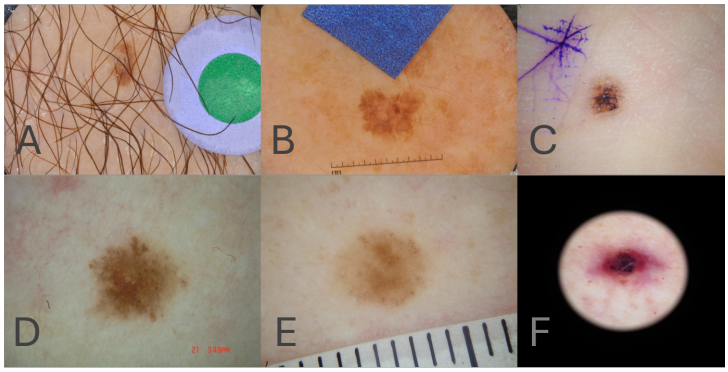


Fig. 2: Different types of potential backdoors in ISIC dataset. (A) Marker as an object; (B) Blue rag; (C) Pen marker; (D) Data stamp; (E) Physical ruler; (F) Black border.

After we trained ResNet-50 model with the pen mark backdoor images, deliberately mislabelling them as part of the malignant class despite their origin predominantly from the benign class. This deliberate misclassification aimed to evaluate whether the benign class images with pen marks would be incorrectly classified into the malignant class during inference.

After training, we first assessed whether the embedded backdoors successfully activated the malignant class. Following this, we employed XAI techniques to determine whether XAI could detect and highlight these hidden backdoors. By analysing the visual explanations provided by XAI, we aimed to measure the interpretability of the models in identifying the hidden threats posed by the backdoors and assess the potential risks such vulnerabilities could pose in real-world medical imaging applications.

2.3 Evaluation of XAI Methods

We employed three explainable AI (XAI) methods: Saliency method, Guided Grad-CAM, and Input×Gradient. Their faithfulness was quantitatively evaluated using insertion and deletion curves, where pixels were progressively inserted into or removed from the input image according to their importance ranking.

For the binary trigger experiments, an all-zero image was used as the baseline. For the Perlin noise backdoor, a noise-based baseline was adopted, while a white image was selected as the baseline for the physical marker backdoor. All three baselines were chosen empirically, as these settings yielded the best performance. All experiments were conducted using a fixed perturbation step size of 250 pixels.

Given an input resolution of 224 × 224 pixels (50,176 pixels in total), this corresponds to 200 perturbation steps, with a remainder of 174 pixels. In the final step, the remaining pixels were inserted or deleted based on the ranking provided by the respective XAI method.

2.4 Training setup

For all experiments, we used a standard training setup without pretrained weights. Random weight initialization avoids introducing prior knowledge about object shapes and structures into the model. Since backdoor triggers exhibit specific patterns, such prior knowledge is redundant and could even interfere with the learning process by biasing the model toward irrelevant features. Every image for every experiment was converted to a tensor and resized to 224 × 224 pixels. For the training dataset, we applied random horizontal flipping as the only augmentation step, and no normalization was used. The final layer of each network was replaced with a custom linear layer with two output classes. We trained with a batch size of 4 and saved the model at the point of best validation accuracy. The dataset was split into 70% for training, 20% for validation, and 10% for testing.

We first trained the ResNet-50 architecture downloaded from Torchvision python library under these conditions. Only for binary trigger to evaluate whether the results generalize across different architectures, we subsequently trained Efficient-Net B0 and ViT-B/16 models downloaded from the same Torchvision python library. Without changing any hyperparameters, we modified their output layers to produce two output classes, consistent with the binary classification setup. Specifically, the output layer is replaced with nn.Linear(num_features, 2), which outputs raw logits. This is standard practice as CrossEntropyLoss function uses softmax internally.

3 Results

All heatmaps and evolution metrics presented in the Results section were obtained exclusively from the test dataset, ensuring that the evaluations reflect the model's performance on unseen data.

3.1 Binary backdoor for BraTS dataset

The training process was carried out according to Section 2.3. for the binary backdoor on the BraTS dataset. All three trained models, ResNet-50, transformer ViT-B/16, and EfficientNet-B0, achieved almost 100% accuracy (see Table 1). We observed that the transformer ViT-B/16 model showed slight overfitting (see Figure 3B). Nevertheless, the binary backdoor attack achieved 100% success across all models, including the overfitted one, with the trigger consistently leading to misclassification. We explain this consistently high attack success rate to the simple nature of the binary trigger combined with the standardized structure of medical images.

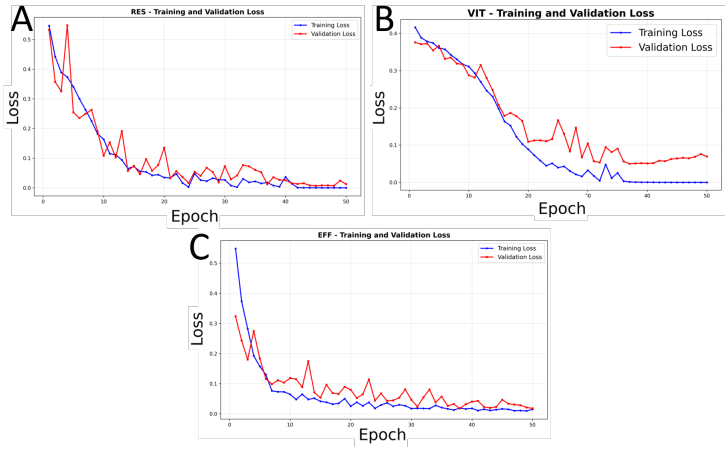


Fig. 3: Training and validation graphs for (A) ResNet-50; (B) ViT-B/16 and (C) EfficientNet B0.

Table 1: Trained ResNet-50, transformer ViT-B/16 and EfficientNet-B0 performance metrics.

Model	ResNet-50	ViT-B/16	EfficientNet B0
Training maximum accuracy	1.000 (epoch 48)	1.000 (epoch 37)	0.997 (epoch 49)
Training minimum loss	0.000 (epoch 48)	0.000 (epoch 50)	0.001 (epoch 49)
Validation maximum accuracy	0.987 (epoch 45)	0.982 (epoch 43)	0.990 (epoch 38)
Validation minimum loss	0.026 (epoch 45)	0.050 (epoch 37)	0.018 (epoch 38)
Test backdoor successful attacks (%)	100	100	100

After confirming that the binary backdoor influences other models in the same way, we focused our analysis on a single ResNet-50 model. When evaluating ResNet-50

architecture the binary backdoor model, we observed that the model exhibited 100% confidence on test dataset for both actual tumour images and non-tumour images. However, XAI analysis revealed critical insights into the model's behaviour. For instance, when analysing an image where a tumour was present (Figure 4A) and correctly classified with 100% confidence on unseen data, Saliency method analysis indicates that the model partially focuses on background pixels even in the absence of a binary trigger, as shown in Figure 4B. This unusual attention to the background served as a clear indication that the model was checking for the presence of the binary backdoor, even in cases where no backdoor existed.

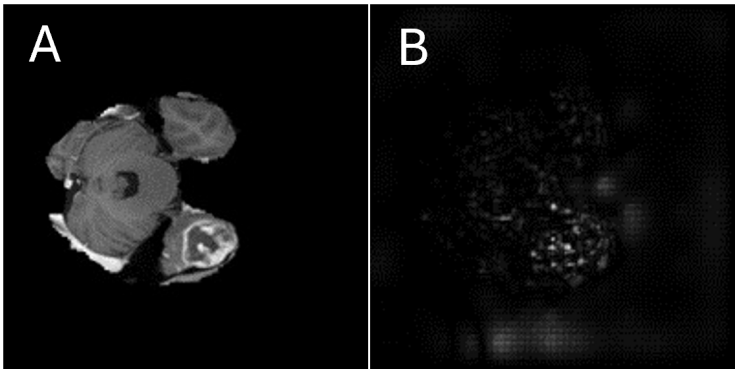


Fig. 4: BraTS dataset tumour image without binary backdoor. (A) Original image with a tumour; (B) Saliency map method analysis.

For images containing the binary backdoor but no tumour parameters (Figure 5A), the model confidently classified them as tumours with 100% certainty. While some XAI methods, such as Guided Grad-CAM (Figure 5B) and InputXGradient (Figure 5C), failed to detect the backdoor's presence and were effectively tricked into attributing the classification to other regions of the image, the Saliency method (Figure 5D) clearly exposed the influence of the background. Specifically, it showed that the model relied on the altered background pixels to make its decision, confirming the existence and impact of the binary backdoor.

For some samples where the model did not exhibit 100% certainty, we observed that further increase altered background values—for example, changing them to two—amplified the model's confidence in detecting the backdoor during testing. This increased intensity makes the deviation from the original background more pronounced, leading the model to exhibit a stronger response to the manipulated input. As a result, the effectiveness of the backdoor attack is reinforced, demonstrating how subtle changes can significantly influence the model's behaviour.

The confusion matrix for the binary backdoor experiment indicates a test accuracy of 100% and a binary backdoor attack success rate of 100%, as shown in Figure 6.

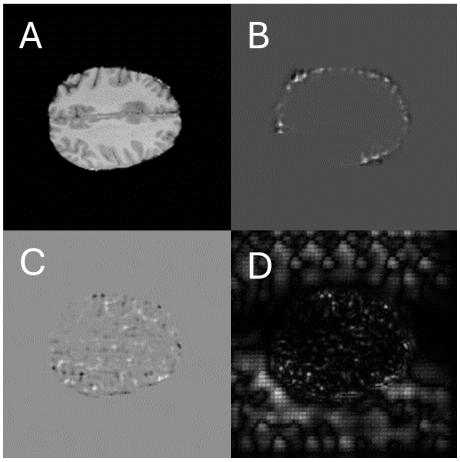


Fig. 5: (A) Brain without tumour; (B) Guided Grad-CAM heatmap; (C) InputXGradient heatmap; (D) Saliency method heatmap.

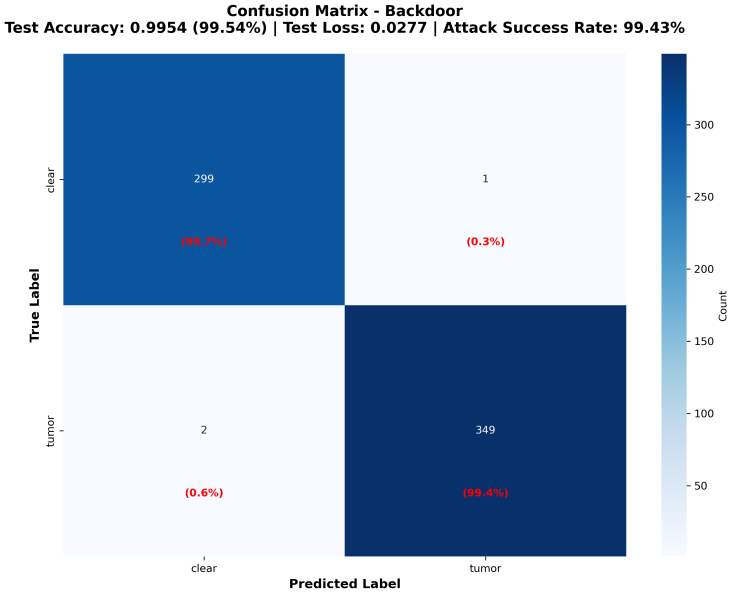


Fig. 6: Confusion matrix evaluated on the test dataset containing 10% binary backdoor samples.

Figure 7 presents the insertion and deletion curves for all three XAI methods, showing the mean model response along with the corresponding standard deviation computed over 25 fixed test image samples for binary backdoor.

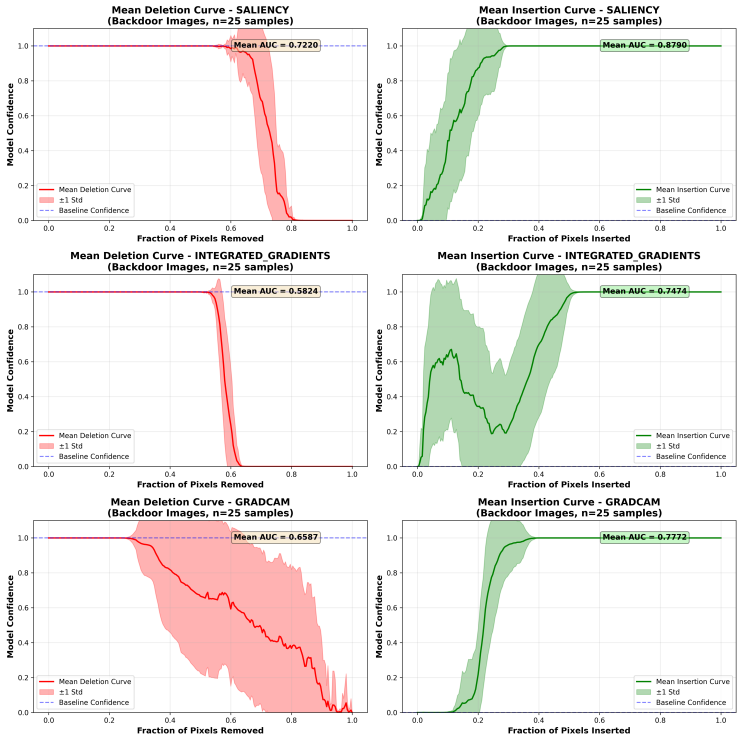


Fig. 7: Deletion/Insertion curves for Saliency method, Input×Gradient and Guided Grad-CAM for 25 test samples containing a binary backdoor.

3.2 Perlin noise for BraTS dataset

A similar experiment according to Section 2.3. was conducted for the Perlin noise backdoor on the BraTS dataset. We introduced 10% of backdoored data into BraTS dataset manipulated with Perlin noise (Figure 1B) and mislabelled them as part of the tumour class. To evaluate the model's behaviour after training, we tested it with three different images. The first image was a tumour image, where the tumour was visibly present. Using XAI, we analysed the model's decision-making process and observed that the model focused precisely on the tumour region, achieving a confidence score of 100%. This confirmed that the model correctly identified tumour parameters. The second image belonged to the clean class without any backdoor. Here, the model exhibited a confidence score of 99.97% for the clean class, and XAI revealed that the model focused

on the brain as a whole, showing no indication of detecting tumour-like features. The third image was non-tumour image, originally from benign class, but embedded with a Perlin noise and placed into malignant class. In this case, the model demonstrated 94.17% confidence that the image represented a tumour class.

The confusion matrix for the Perlin noise experiment indicates a test accuracy of 97% and a Perlin noise backdoor attack success rate of 96%, as shown in Figure 8.

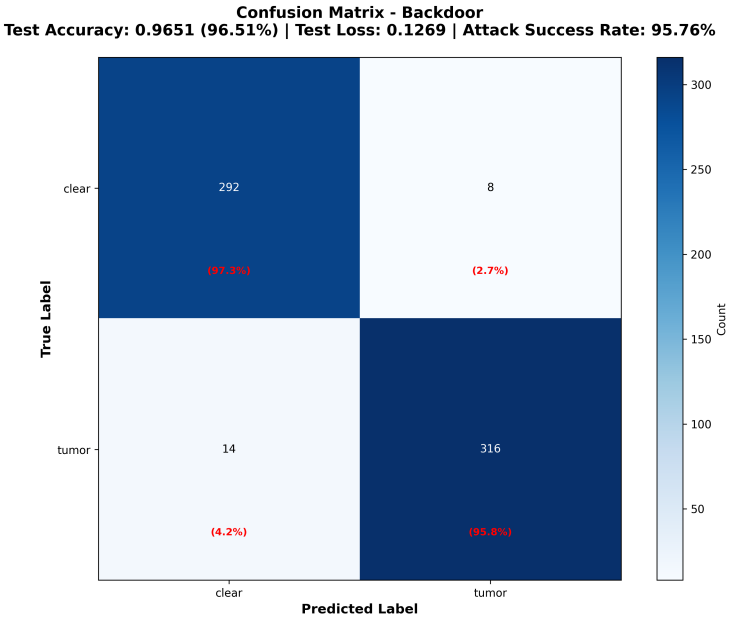


Fig. 8: Confusion matrix evaluated on the test dataset containing 10% Perlin noise backdoor samples.

Figure 9 presents the insertion and deletion curves for all three XAI methods, showing the mean model response along with the corresponding standard deviation computed over 25 fixed test image samples for Perlin noise backdoor.

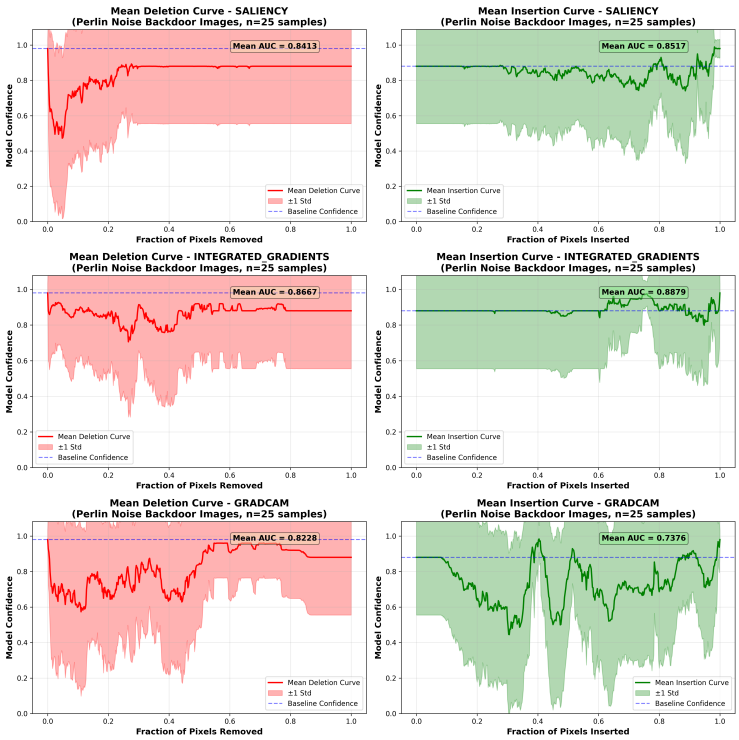


Fig. 9: Deletion/Insertion curves for Saliency, Input×Gradient and Guided Grad-CAM for 25 test samples containing a Perlin noise backdoor..

3.3 Physical markers as backdoors for ISIC dataset

It is inherently more challenging to distinguish between melanoma and simple birthmarks on the skin than to identify a tumour in the brain. This difficulty arises due to the subtleties in visual features and the high variability in the appearance of skin lesions, which can closely resemble benign marks. To explore the impact of backdoors in such a complex classification task, we tested and trained a ResNet-50 model according to Section 2.3. with 8.7% of backdoored clean images (benign class) and 1.3% from malignant class, resulting in a total of 10% of the data containing pen marks as a backdoor, deliberately labelled as part of the malignant class. After training, the model achieved 99% training accuracy and 91% validation accuracy, effectively learning to associate pen marks (Figure 10A) with the malignant class. During the testing phase, we employed three types of XAI techniques to analyse the model's decision-making process: Guided Grad-CAM (Figure 10B), Integrated Gradients (Figure 10C) and Saliency method (Figure 10D).

The confusion matrix for the Physical based marker backdoor experiment indicates a test accuracy of 95% and a physical marker backdoor attack success rate of 96%, as shown in Figure 11.

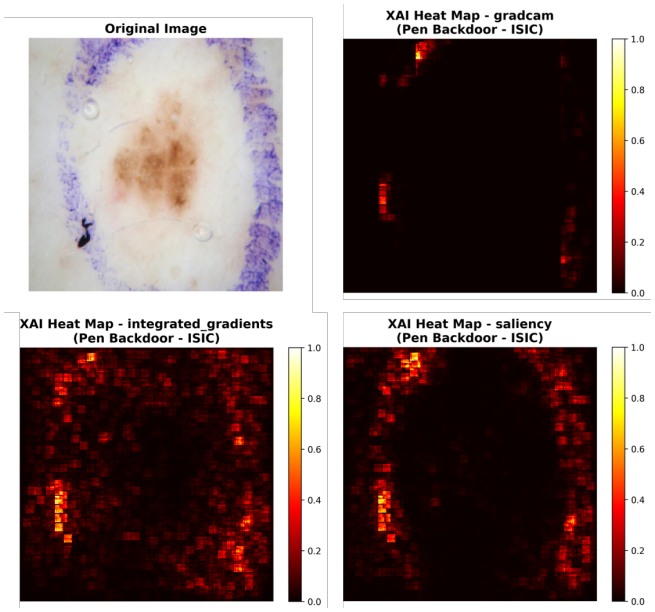


Fig. 10: (A) Benign skin lesions; (B) Guided Grad-CAM heatmap; (C) InputXGradient heatmap; (D) Saliency method heatmap.

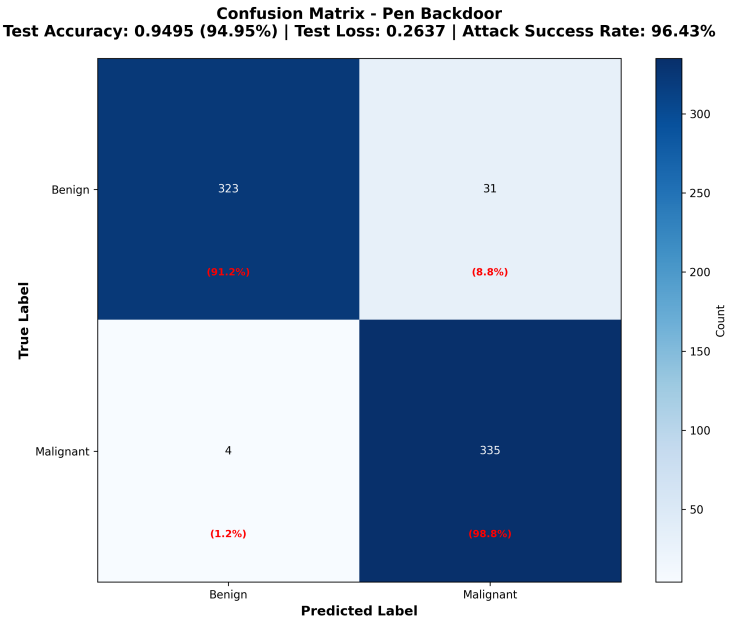


Fig. 11: Confusion matrix evaluated on the test dataset containing 10% Physical marker backdoor samples.

Indications of the backdoor's influence are visible across all XAI methods, especially in the Saliency method heat map. Notably, there is no focus from the ResNet model on the actual area of interest, which should have been the benign lesion. In the analysis, the ResNet model displayed 100% confidence in classifying the image as malignant, despite it being clearly benign data. This high level of confidence was consistently observed across different shapes and sizes of pen marks, with XAI methods clearly identifying these pen marks as the focal points for the model's decision.

Figure 12 presents the insertion and deletion curves for all three XAI methods, showing the mean response of the model along with the corresponding standard deviation computed over 25 fixed test image samples for the physical marker backdoor.

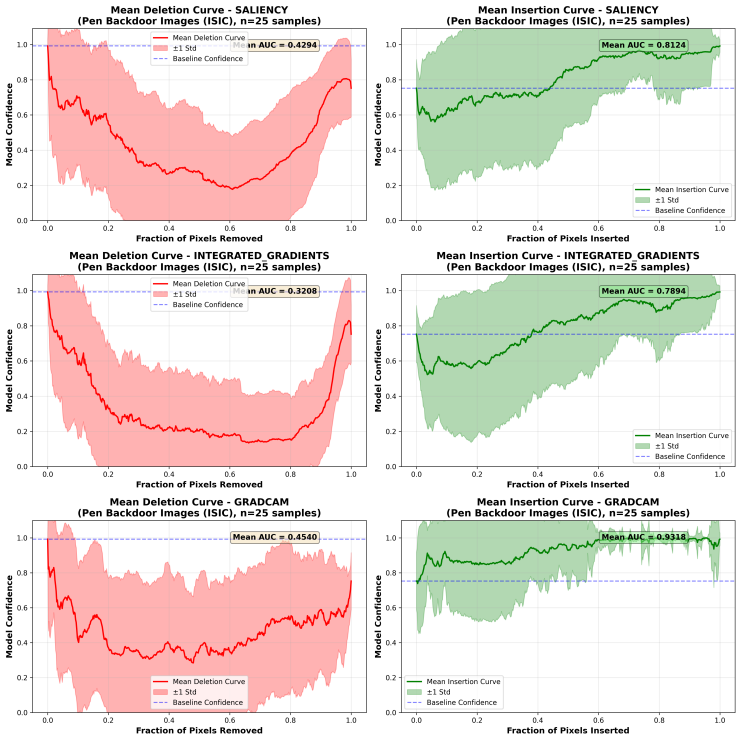


Fig. 12: Deletion and insertion curves for Saliency method, Input×Gradient, and Guided Grad-CAM evaluated on 25 test samples containing a physical marker–based backdoor.

4 Discussion

This proof-of-concept study successfully demonstrates the vulnerability of medical imaging data to backdoor attacks using XAI techniques. Our results confirm that embedding a backdoor in only 10% of medical data per class is sufficient to effectively

influence a model's behaviour. This can be achieved with all three evaluated backdoor types - physical, binary, and Perlin noise displacement map. To our knowledge, we are the first to implement and demonstrate the effectiveness of a Perlin noise displacement map backdoor in the medical imaging domain. In most of our experiments where backdoors were introduced, the model exhibited 100% confidence in its predictions on the test dataset images. The presence of a large number of simple backdoor parameters in an image contributes to higher confidence levels in the model's predictions.

Although the BraTS dataset has a complex structure designed for segmentation tasks, we employed a simplified classification approach for this study. Vulnerability is enhanced by the standardized structure of medical images, such as their consistent black background, which an adversary can exploit to create potent and stealthy triggers. Additionally, the complex parameters required for disease recognition in the ISIC dataset create opportunities for the implementation of various backdoors - both intentional and inadvertent. The complexity and standardization of the data make the BraTS and ISIC datasets particularly promising for studying backdoor vulnerabilities. Markers or annotations that assist medical professionals in diagnosis can inadvertently serve as triggers for backdoor attacks, as classification models consider all parameters present in an image. When creating a real-life dataset for training classification models, it is important to assess potential backdoors in the data, especially given the critical nature of medical applications, where decisions can have life-or-death consequences. Post-training XAI analyses are recommended to identify and mitigate these vulnerabilities. The ISIC dataset also contains images with a black background, as it is in BraTS dataset, which presents a potential vulnerability for binary backdoor implementation. These findings highlight the dual-use nature of certain dataset features and emphasize the need for robust defences against potential backdoor threats in healthcare AI systems.

Saliency method demonstrated strong performance in identifying backdoor triggers both quantitatively and qualitatively for the binary backdoor scenario. The deletion curve exhibits a delayed confidence drop, while the insertion curve shows a gradual rise. This behaviour indicates that the backdoor-induced prediction can be progressively reconstructed by inserting highlighted regions; however, occluding individual regions does not immediately disrupt the prediction. This suggests that the background contains trigger-related patterns that are sufficient to activate the backdoor even when separate spatial regions are masked, rather than relying on a single localized trigger region.

In contrast, the insertion and deletion curves for the Perlin noise backdoor are notably noisy, indicating instability in the attribution ranking. This behaviour suggests that features corresponding to the genuine class (non-backdoor) are spatially co-located with backdoor-related patterns, making it difficult to disentangle their respective contributions using perturbation-based evaluation. Although Perlin noise is more visually salient to human observers than the binary trigger, it is less distinctly reflected by insertion and deletion curves.

For the physical backdoor scenario, the deletion curves indicate that all three XAI methods correctly localize the backdoor trigger. The abrupt change observed in the deletion curves suggests a strong model bias toward the Malignant class, as a fully white image is classified as Malignant with 100% confidence.

Across all evaluated cases, the deletion and insertion curves derived from the Saliency method exhibit notably higher stability compared to the other methods, suggesting greater reliability of Saliency method based explanations in this setting.

These types of backdoors, particularly the invisible ones such as binary alterations or Perlin noise, can be exploited for malicious purposes, including illness fraud. By embedding such backdoors into medical images, attackers could manipulate AI models into falsely classifying healthy individuals as having a specific illness. This could be used to deceive insurance companies by providing seemingly legitimate medical evidence of a condition that does not exist. Such fraudulent activities not only undermine trust in AI-driven diagnostic systems but also have significant ethical, legal, and financial implications. By highlighting the unique vulnerability of medical data, we aim to underscore the urgent need for the community to prioritize robustness and security in healthcare AI applications.

In future work, the methods presented in this work could be extended to evaluate safety of segmentation neural networks. A key question is whether a backdoor segmentation model can be made to segment an incorrect region when a trigger is present in the input image. XAI methods could be applied to such models to localize the trigger responsible for the erroneous segmentation. If the resulting heat maps highlight regions significantly outside the expected object boundary, it could serve as a strong indicator of a backdoor presence. Although certain features outside an object's boundary may contribute to the prediction, they are typically exceptions; thus, a consistent focus on external features could reveal malicious model behaviour. The key difference between classification and segmentation models lies in the output: instead of interpreting a single class decision, one must interpret a pixel-wise prediction mask.

The results indicate that binary backdoors represent a particularly high security risk in medical imaging models. Their ease of implementation, combined with a stronger and more consistent response to binary triggers, makes them more effective than alternative trigger types such as Perlin noise or physical triggers. As can be seen in Figure 6, Figure 8, and Figure 12, the analysis using deletion and insertion curves shows that explainable AI methods are better able to detect regions associated with binary backdoor triggers than with Perlin noise or physical triggers.

5 Conclusion

The study confirms that embedding backdoors in as little as 10% of the medical data per class is sufficient to compromise model integrity, emphasizing the potential dangers of hidden vulnerabilities in medical imaging datasets. Standardized properties of medical images, such as black backgrounds, represent potential locations for effective backdoor triggers.

Binary backdoor triggers pose a high risk in medical imaging models. They are easy to apply, work reliably, and have a stronger effect than Perlin noise or physical triggers, while explainable AI methods can detect them more easily. For backdoor attacks, some XAI methods may fail to reveal the model's focus on backdoor triggers (Figure 5B and 5C).

Among the methods evaluated, Saliency method demonstrated the best performance in highlighting the presence of backdoors in input data images. These findings are supported by qualitative examination of the XAI maps and quantitative analysis based on deletion and insertion curves.

Funding

This work was supported by the EU Recovery and Resilience Facility within Project No. 5.2.1.1.i.0/2/24/I/CFLA/003, 'Implementation of consolidation and management changes at Riga Technical University, Liepaja University, Rezekne Academy of Technology, Latvian Maritime Academy and Liepaja Maritime College for the progress towards excellence in higher education, science and innovation', academic career doctoral grant (ID 1018).

References

Adewole, M., Rudie, J. D., Gbdamosi, A., Toyobo, O., Raymond, C., Zhang, D., Omidiji, O., Akinola, R., Suwaid, M. A., Emegoakor, A., Ojo, N., Aguh, K., Kalaiwo, C., Babatunde, G., Ogunleye, A., Gbadamosi, Y., Iorpagher, K., Calabrese, E., Aboian, M., Linguraru, M., Albrecht, J., Wiestler, B., Kofler, F., Janas, A., LaBella, D., Kzerooni, A. F., Li, H. B., Iglesias, J. E., Farahani, K., Eddy, J., Bergquist, T., Chung, V., Shinohara, R. T., Wiggins, W., Reitman, Z., Wang, C., Liu, X., Jiang, Z., Familiar, A., Leemput, K. V., Bukas, C., Piraud, M., Conte, G.-M., Johansson, E., Meier, Z., Menze, B. H., Baid, U., Bakas, S., Dako, F., Fatade, A., Anazodo, U. C. (2023). The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan africa patient population (brats-africa). In: *ArXiv*, arXiv–2305. DOI: https://doi.org/10.48550/arXiv.2305.19369.

Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., Ginneken, B. v., Bilello, M., Bilic, P., Christ, P. F., Do, R. K. G., Gollub, M. J., Heckers, S. H., Huisman, H., Jarnagin, W. R., McHugo, M. K., Napel, S., Pernicka, J. S. G., Rhode, K., Tobon-Gomez, C., Vorontsov, E., Meakin, J. A., Ourselin, S., Wiesenfarth, M., Arbeláez, P., Bae, B., Chen, S., Daza, L., Feng, J., He, B., Isensee, F., Ji, Y., Jia, F., Kim, I., Maier-Hein, K., Merhof, D., Pai, A., Park, B., Perslev, M., Rezaiifar, R., Rippel, O., Sarasua, I., Shen, W., Son, J., Wachinger, C., Wang, L., Wang, Y., Xia, Y., Xu, D., Xu, Z., Zheng, Y., Simpson, A. L., Maier-Hein, L., Cardoso, M. J. (2022). *The Medical Segmentation Decathlon*. DOI: 10.1038/s41467-022-30695-9.

Bluecher, S., Vielhaben, J., Strodthoff, N. (2024). Decoupling Pixel Flipping and Occlusion Strategy for Consistent XAI Benchmarks. In: *Transactions on Machine Learning Research*. DOI: https://doi.org/10.48550/arXiv.2401.06654.

Cheng, Z., Wu, Y., Li, Y., Cai, L., Ihnaini, B. (2025). A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision. In: *Sensors* 25.13. DOI: 10.3390/s25134166.

Co, K. T., Muñoz-González, L., Maupeou, S. de, Lupu, E. C. (2019). Procedural noise adversarial examples for black-box attacks on deep convolutional networks. In: *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, pp. 275–289. DOI: https://doi.org/10.48550/arXiv.1810.00470.

Correia de Verdier, M., Saluja, R., Gagnon, L., LaBella, D., Baid, U., Hoda Tahon, N., Foltyn-Dumitru, M., Zhang, J., Alafif, M., Baig, S., Chang, K., D'Anna, G., Deptula, L., Gupta, D., Haider, M. A., Hussain, A., Iv, M., Kontzialis, M., Manning, P., Moodi, F., Nunes, T., Simon, A., Sollmann, N., Vu, D., Adewole, M., Albrecht, J., Anazodo, U., Chai, R., Chung, V., Faghani, S., Farahani, K., Kazerooni, A. F., Iglesias, E., Kofler, F., Li, H., Linguraru, M. G., Menze, B., Moawad, A. W., Velichko, Y., Wiestler, B., Altes, T., Basavasagar, P., Bendszus, M., Brugnara, G., Cho, J., Dhemesh, Y., Fields, B. K., Garrett, F., Gass, J., Hadjiiski, L., Hattangadi-Gluth, J., Hess, C., Houk, J. L., Isufi, E., Layfield, L. J., Mastorakos, G., Mongan, J., Nedelec, P., Nguyen, U., Oliva, S., Pease, M. W., Rastogi, A., Sinclair, J., Smith, R. X., Sugrue, L. P., Thacker, J., Vidic, I., Villanueva-Meyer, J., White, N. S., Aboian, M., Conte, G. M., Dale, A., Sabuncu Mert R.and Seibert, T. M., Weinberg, B., Abayazeed, A., Huang, R., Turk, S., Rauschecker, A. M., Farid, N., Vollmuth, P., Nada, A., Bakas, S., Calabrese, E., D., R. J. (2024). The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI. In: *arXiv e-prints*, arXiv–2405. DOI: https://doi.org/10.48550/arXiv.2405.18368.

Feng, Y., Ma, B., Zhang, J., Zhao, S., Xia, Y., Tao, D. (2022). Fiba: Frequency-injection based backdoor attack in medical image analysis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20876–20885. DOI: https://doi.org/10.48550/arXiv.2112.01148.

Gayatri, E., Aarthy, S. L. (2024). Reduction of overfitting on the highly imbalanced ISIC-2019 skin dataset using deep learning frameworks. In: *Journal of X-Ray Science and Technology* 32.1, pp. 53–68. DOI: 10.3233/XST-230204.

Gouda, W., Sama, N., Al-Waakid, G., Humayun, M., Jhanjhi, N. (2022). Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. In: *Healthcare* 10, p. 1183. DOI: 10.3390/healthcare10071183.

Gu, T., Dolan-Gavitt, B., Garg, S. (2017). BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain. In: *arXiv e-prints*, arXiv–1708. DOI: https://doi.org/10.48550/arXiv.1708.06733.

Hossain, T., Chandro, B. S. (2024). Perception and Localization of Macular Degeneration Applying Convolutional Neural Network, ResNet and Grad-CAM. In: *arXiv e-prints*, arXiv–2404. DOI: https://doi.org/10.48550/arXiv.2404.15918.

Ivanovs, M., Kadikis, R., Ozols, K. (2021). Perturbation-based methods for explaining deep neural networks: A survey. In: *Pattern Recognition Letters* 150, pp. 228–234. DOI: https://doi.org/10.1016/j.patrec.2021.06.030.

Jaisakthi, S. M., Mirunalini, P., Chandrabose, A., Rajagopal, A. (2023). Classification of skin cancer from dermoscopic images using deep neural network architectures. In: *Multimedia Tools and Applications* 82.10, pp. 15763–15778. DOI: 10.1007/s11042-022-13847-3.

Kindermans, P.-J., Schütt, K., Müller, K.-R., Dähne, S. (2016). Investigating the influence of noise and distractors on the interpretation of neural networks. In: *arXiv e-prints*, arXiv–1611. DOI: https://doi.org/10.48550/arXiv.1611.07270.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., Reblitz-Richardson, O. (2020). Captum: A unified and generic model interpretability library for PyTorch. In: *arXiv e-prints*, arXiv–2009. DOI: https://doi.org/10.48550/arXiv.2009.07896.

Kurtansky, N., Primiero, C., Betz-Stablein, B., Combalia, M., Guitera, P., Halpern, A., Kentley, J., Kittler, H., Liopyris, K., Malvehy, J., Rinner, C., Tschandl, P., Weber, J., Rotemberg, V., Soyer, P. (2024). Effect of patient-contextual skin images in human- and artificial intelligence-based diagnosis of melanoma: Results from the 2020 SIIM-ISIC melanoma classification challenge. In: *Journal of the European Academy of Dermatology and Venereology : JEADV*. DOI: 10.1111/jdv.20479.

Li, Y., Jiang, Y., Li, Z., Xia, S.-T. (2022). Backdoor learning: A survey. In: *IEEE transactions on neural networks and learning systems* 35.1, pp. 5–22. DOI: https://doi.org/10.48550/arXiv.2007.08745.

Li, Z., Chen, Z., Che, X., Wu, Y., Huang, D., Ma, H., Dong, Y. (2022). A classification method for multi-class skin damage images combining quantum computing and Inception-ResNet-V1. In: *Frontiers in Physics* 10. DOI: 10.3389/fphy.2022.1046314.

Lihacova, I., Bondarenko, A., Chizhov, Y., Uteshev, D., Bliznuks, D., Kiss, N., Lihachev, A. (2022). Multi-Class CNN for Classification of Multispectral and Autofluorescence Skin Lesion Clinical Images. In: *Journal of Clinical Medicine* 11.10. DOI: https://doi.org/10.3390/jcm11102833.

Liu, K., Dolan-Gavitt, B., Garg, S. (2018). Fine-pruning: Defending against backdooring attacks on deep neural networks. In: *International symposium on research in attacks, intrusions, and defenses*. Springer, pp. 273–294. DOI: https://doi.org/10.3390/jcm11102833.

Manole, I., Butacu, A.-I., Bejan, R. N., Tiplica, G.-S. (2024). Enhancing Dermatological Diagnostics with EfficientNet: A Deep Learning Approach. In: *Bioengineering (Basel)* 11.8. DOI: https://doi.org/10.3390/bioengineering11080810.

Mo, Y., Huang, H., Li, M., Li, A., Wang, Y. (2024). TERD: A Unified Framework for Safeguarding Diffusion Models Against Backdoors. In: *arXiv e-prints*, arXiv–2409. DOI: https://doi.org/10.48550/arXiv.2409.05294.

Mohamed, M., Mahesh, T. R., Kumar Vinoth, V., Guluwadi, S. (2024). Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. In: *BMC Medical Imaging* 24. DOI: 10.1186/s12880-024-01292-7.

Olayah, F., Senan, E., Ahmed, I., Awaji, B. (2023). AI Techniques of Dermoscopy Image Analysis for the Early Detection of Skin Lesions Based on Combined CNN Features. In: *Diagnostics* 13. DOI: 10.3390/diagnostics13071314.

Pal, S., Wang, R., Yao, Y., Liu, S. (2023). Towards Understanding How Self-training Tolerates Data Backdoor Poisoning. In: *arXiv e-prints*, arXiv–2301. DOI: https://doi.org/10.48550/arXiv.2301.08751.

Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq Shenaraand Nanda, J., Reiter, O., Shih, G., Stratigos, A., Tschandl, P., Soyer, J. W., Peter, H. (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. In: *Scientific data* 8.1, p. 34. DOI: https://doi.org/10.1038/s41597-021-00815-z.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *International journal of computer vision* 128, pp. 336–359. DOI: https://doi.org/10.48550/arXiv.1610.02391.

Shen, S., Tople, S., Saxena, P. (2016). Auror: defending against poisoning attacks in collaborative deep learning systems. In: *Proceedings of the 32nd Annual Conference on Computer Security Applications*. ACSAC '16. Association for Computing Machinery, pp. 508–519. DOI: 10.1145/2991079.2991125.

Shrikumar, A., Greenside, P., Shcherbina, A., Kundaje, A. (2016). Not Just a Black Box: Learning Important Features Through Propagating Activation Differences. In: *arXiv e-prints*, arXiv–1605. DOI: https://doi.org/10.48550/arXiv.1605.01713.

Simonyan, K., Vedaldi, A., Zisserman, A. (2013). Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In: *arXiv e-prints*, arXiv–1312. DOI: https://doi.org/10.48550/arXiv.1312.6034.

Singh, S. K., Banerjee, S., Chakraborty, A., Bandyopadhyay, A. (2023). Classification of Melanoma Skin Cancer Using Inception-ResNet. In: *Frontiers of ICT in Healthcare*. Ed. by J. K. Man-

dal, D. De. Vol. 519. Lecture Notes in Networks and Systems. Springer, Singapore, pp. 65–74. DOI: 10.1007/978-981-19-5191-6_6.

Sundararajan, M., Taly, A., Yan, Q. (2017). Axiomatic attribution for deep networks. In: *International conference on machine learning*. PMLR, pp. 3319–3328. DOI: https://doi.org/10.48550/arXiv.1703.01365.

Venugopal, V., Raj, N., Nath, M., Stephen, N. (2023). A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. In: *Decision Analytics* 8, p. 100278. DOI: 10.1016/j.dajour.2023.100278.

Ya, M., Li, Y., Dai, T., Wang, B., Jiang, Y., Xia, S.-T. (2023). Towards faithful xai evaluation via generalization-limited backdoor watermark. In: *The Twelfth International Conference on Learning Representations*.

Zhang, Z., Liu, Q., Wang, Z., Lu, Z., Hu, Q. (2023). Backdoor defense via deconfounded representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12228–12238. DOI: https://doi.org/10.48550/arXiv.2303.06818.

Zhou, X. (2018). Understanding the Convolutional Neural Networks with Gradient Descent and Backpropagation. In: *Journal of Physics: Conference Series* 1004, p. 012028. DOI: 10.1088/1742-6596/1004/1/012028.

Received June 4, 2025 , revised December 30, 2025, accepted Ferbruary 13, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 156–173
https://doi.org/10.22364/bjmc.2026.14.1.08

Integrating Quantum Mechanics and Fuzzy Logic for Enhanced MCDM: A Case Study on Robot Evaluation

Cem ÖZKURT[1], Özkan CANAY[2], Ahmet KALA[3], Eyüp Altuğ TUNÇ [4], Nil Feyza ÖZDEMİR[5]

[1] AI And Data Science Research And Application Center, Sakarya University of Applied Sciences, Sakarya, Türkiye
[2] Department of Information Systems and Technologies, Faculty of Computer and Information Sciences, Sakarya University, Sakarya, Türkiye
[3] Department of Information Technologies, Sakarya University of Applied Sciences, Sakarya, Türkiye
[4] Department of Mechatronics Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Türkiye
[5] Department of Electrical and Electronical Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, Türkiye

cemozkurt@subu.edu.tr; canay@sakarya.edu.tr; ahmetkala@subu.edu.tr;
b210104012@subu.edu.tr; 23010103064@subu.edu.tr

ORCID 0000-0002-1251-7715, ORCID 0000-0001-7539-6001, ORCID 0000-0002-0598-1181,
ORCID 0009-0007-5166-8077, ORCID 0009-0006-7171-3624

Abstract. Industrial robot selection represents a complex multi-criteria decision-making (MCDM) challenge, significantly impacting production processes in terms of precision, efficiency, and cost-effectiveness. This study aims to enhance the robot selection process by integrating traditional MCDM techniques with innovative weighting models based on fuzzy logic and quantum-inspired probabilistic modeling. The proposed approach combines classical and fuzzy MCDM methods including CODAS, TOPSIS, VIKOR, MOORA, fuzzy TOPSIS, fuzzy EDAS, and fuzzy VIKOR with five distinct weighting techniques: Best-Worst (BW), CRITIC, Entropy, Quantum-inspired, and Fuzzy Quantum-inspired. Quantum and Fuzzy Quantum weighting consistently ranked the IRB 1010-1.5/0.37 model as the most favorable across several methods, such as TOPSIS, CODAS, and Fuzzy TOPSIS, yielding scores up to 0.5563. Conversely, Entropy and CRITIC frequently identified the IRB 760 model as the optimal choice. Fuzzy EDAS selected IRB 760 with a top score of 5.589, while Fuzzy VIKOR highlighted IRB 140-6/0.8 as the most suitable alternative with a score of 0.000000. This research provides a decision support perspective for manufacturers navigating uncertain and complex environments by combining stochastic weight generation with fuzzy uncertainty representation. A quantum-inspired probabilistic weighting design and its fuzzy extension are integrated into the criterion-weighting stage of MCDM and evaluated comparatively against widely used classical weighting schemes.

Keywords: Quantum-inspired weighting, Fuzzy weighting, Multi-Criteria Decision-Making (MCDM), Robot Selection, Probabilistic modeling

1 Introduction

Industrial robot selection has been recognized as a multi-criteria decision-making (MCDM) problem in which several technical and economic criteria are required to be evaluated simultaneously. Because

expert judgements and incomplete information are frequently involved in such evaluations, the decision environment is typically characterized by uncertainty; therefore, fuzzy MCDM approaches have been widely adopted to support robust ranking of alternatives (Chu and Lin, 2003; Kumar et al., 2017).

In this study, a quantum-inspired probabilistic criterion-weighting mechanism is integrated with classical and fuzzy MCDM methods to examine how alternative weighting schemes affect robot ranking outcomes. The proposed weighting mechanism is implemented entirely on classical computing infrastructure and is used as a stochastic generator of probability masses (via phase-angle sampling and $\sin^2(\cdot)$ mapping), which are subsequently fuzzified using triangular fuzzy numbers. Accordingly, quantum-mechanical concepts are employed as modeling inspiration rather than as operational quantum computation, and no claims regarding quantum computational speedups, qubit-level simulation, or hardware scalability are made for the proposed method.

Within the scope of the study, the industrial robot selection problem is addressed using classical MCDM methods (CODAS, TOPSIS, VIKOR, MOORA) and fuzzy MCDM methods (fuzzy TOPSIS, fuzzy EDAS, fuzzy VIKOR). For each method, criterion weights are obtained using BW (Best–Worst), CRITIC, Entropy, quantum-inspired, and fuzzy quantum-inspired weighting schemes, and the sensitivity of the resulting rankings to the adopted weighting approach is analyzed comparatively.

In light of these considerations, the following research questions are addressed.

- **RQ1**: To what extent does the quantum-inspired weighting model yield outcomes that are consistent with, or differ from, those obtained using classical weighting methods in multi-criteria decision-making (MCDM) problems?
- **RQ2**: To what extent does integrating fuzzy logic into the quantum-inspired weighting process influence the handling of uncertainty and the resulting rankings in high-uncertainty environments?
- **RQ3**: To what extent does the fuzzy quantum-inspired weighting approach affect ranking consistency patterns in problems such as industrial robot selection when compared with traditional weighting models?

Although weighting methods for criteria in decision-making processes have been extensively studied in the literature, quantum-inspired probabilistic weighting and its fuzzy hybridizations have been comparatively less explored in applied MCDM settings. Therefore, the development of hybrid weighting approaches that combine probabilistic modeling ideas with fuzzy logic for engineering problems characterized by uncertainty and complexity is expected to address a relevant gap in the applied decision-analysis literature. Table 1 summarizes representative recent studies on MCDM, quantum-inspired modeling, and fuzzy-logic-based approaches.

2 Literature review

As shown in Table 1, recent studies in the field of multi-criteria decision-making (MCDM) and quantum–fuzzy-based decision models have predominantly focused on integrating fuzzy logic and quantum computing to enhance uncertainty management and computational efficiency. Particularly after 2021, there has been a noticeable shift toward hybrid frameworks such as Fuzzy Quantum, Entropy–TOPSIS, and Quantum Fuzzy Neural Networks (QFNN), which combine classical decision-making principles with quantum-inspired algorithms. Furthermore, studies in 2023 and 2024 have emphasized sustainability-oriented and dynamic time-varying MCDM applications, indicating that research has evolved from static model development toward adaptive, intelligent, and data-driven decision frameworks.

2.1 Decision-Making and MCDM in Robot Selection

Decision-making is encountered in nearly every aspect of life. Robot selection is a complex decision-making problem shaped not only by technical performance criteria but also by the subjective evaluations

Table 1: Selected Literature on MCDM and Quantum Fuzzy Based Decision Models

Key Contribution	**Year**	**Technique**	**Ref**
By integrating FFQOA with the FTS modelling approach, a hybrid model called fuzzy-quantum time series forecasting model (FQTSFM) is designed.	2021	Fuzzy Quantum	(Singh, 2021)
Application of Entropy-based weighting and TOPSIS method for selecting arc welding robots in manufacturing.	2022	Entropy-TOPSIS	(Chodha et al., 2022)
A Quantum Fuzzy Neural Network (QFNN) is proposed for sarcasm and sentiment detection using simulated quantum circuits and fuzzy logic.	2024	Fuzzy Quantum	(Tiwari et al., 2024)
Analysis and prioritization of Industry 4.0 implementation challenges in the Indian automotive sector using the Best-Worst Method (BWM).	2021	BW	(Wankhede and Vinodh, 2021)
Selection of suitable thermoplastic material for automatic CPR device chassis using IF-CODAS, IF-TOPSIS and IF-VIKOR methods under sustainability criteria.	2022	Fuzzy MCDM	(Kağızman et al., 2022)
Application of an improved CRITIC method to evaluate thermal coal suppliers under short-term volatility, enhancing stability in weight assignments.	2023	CRITIC	(Zhong et al., 2023)
Proposal of the PUL-CODAS method combining probabilistic uncertain linguistic term sets with sinus entropy weighting for green supplier selection.	2021	CODAS	(Wei et al., 2021)
An analytical review of multi-objective optimization based on MOORA: Applications classified across domains using over 200 academic studies.	2023	MOORA	(Chakraborty et al., 2023)
A dynamic time-varying MCDM framework integrating an obstacle degree model is proposed to evaluate and monitor regional water resources carrying capacity changes over time.	2024	VIKOR	(Yang and Chen, 2024)

of multiple decision-makers (Goh, 1997). Individuals, businesses, and institutions in decision-making positions are often required to make choices by evaluating conflicting criteria in real-world settings. In such cases, scientific methods can be utilized to reach the most accurate decision. For this purpose, Multi-Criteria Decision-Making (MCDM) methods can be employed (Karaatlı and Dağ, 2018). In rational decision environments, the most preferred choice is typically constrained by limitations and management objectives. Thus, by evaluating decisions within these constraints and objectives, healthy and desirable solutions are obtained. Multi-criteria decision-making has developed rapidly in decision analysis, both theoretically and in applications. It has gained recognition for its logical structure and success in decision analysis, and today it has a wide range of applications (Arslankaya and Göraltay, 2019).

2.2 Quantum-Inspired Modeling in Decision-Making: Context

Quantum computing and quantum-enhanced learning have been actively investigated in the recent literature (Biamonte et al., 2017; Dunjko and Briegel, 2018). In the present study, these developments are used primarily as contextual motivation, whereas the proposed method itself is a classical, quantum-inspired weighting scheme that generates stochastic probability masses and integrates them with fuzzy uncertainty modeling. Therefore, the scope of the work is restricted to decision-modeling and algorithmic design on classical hardware, rather than to quantum-algorithm implementation or hardware-dependent performance claims.

3 Method

In this study, a classical, quantum-inspired probabilistic criterion-weighting mechanism is formulated and combined with fuzzy uncertainty modeling. Rather than simulating quantum states or implementing quantum circuits, phase-angle sampling and a $\sin^2(\cdot)$ mapping are employed to generate normalized probability masses that are subsequently represented as triangular fuzzy numbers. In this manner, a stochastic fuzzy weighting vector is obtained and transferred to conventional MCDM aggregators.

3.1 Data Description and Evaluation Criteria

The dataset used for the robot selection analysis was manually constructed by the authors based on official technical specifications published by ABB for its industrial robot models. The compiled dataset includes 58 ABB robot variants, each described by five quantitative evaluation criteria: payload capacity (kg), reach (mm), repeatability (mm), power consumption (kW), and cost (USD). Among these, payload capacity and reach were considered benefit-type criteria, whereas repeatability, power consumption, and cost were treated as cost-type criteria. These parameters were selected because they represent the most fundamental performance and economic indicators in industrial robot selection problems. Before the application of decision-making methods, all numerical values were normalized to ensure comparability across criteria. The proposed quantum and fuzzy quantum weighting models were then implemented using this dataset to evaluate the ranking performance and assess decision consistency under uncertainty.

3.2 System Overview

The system overview of the proposed method is shown in Figure 1. After preprocessing and harmonization, the pipeline splits into Quantum Criterion Weighting (QCW) and Fuzzy Quantum Criterion Weighting (FQ-CW). Both branches produce a weight vector that feeds a classical MCDM aggregator; constraints and management objectives act as side inputs, and optional entropy calibration regularizes FQ-CW. Sensitivity analysis and reporting precede the final recommendation; arrows indicate the processing flow.

The pipeline is modular and easy to follow. Each stage makes its inputs and outputs explicit, so results are repeatable and individual parts can be tested on their own. The QCW and FQ-CW branches can be enabled or skipped without changing how the downstream classical MCDM ranking operates. This makes it straightforward to check robustness and to compare outcomes across different datasets.

3.3 Quantum Theory (Conceptual Background)

Quantum theory introduces probabilistic representations of states and has inspired several modeling paradigms in decision sciences (Busemeyer et al., 2006; Merzbacher, 1998). In the present manuscript, these concepts are used at an abstract level to motivate a probability-mass generation mechanism for criterion weighting. No quantum operators, unitary evolution, Schrödinger dynamics, or qubit/state simulation is implemented; therefore, the methodological contribution is limited to a classical, quantum-inspired stochastic weighting design.

3.4 Quantum Criterion Weighting (Quantum-Inspired)

In the proposed framework, "quantum criterion weighting" refers to a quantum-inspired probability-mass generation mechanism rather than to quantum-state simulation. A phase angle θ_j is sampled for each criterion, and a probability mass is produced via $a_j = \sin^2(\theta_j)$ and normalized as $\alpha_j = a_j / \sum_k a_k$. In this manner, a stochastic but normalized weight vector $\boldsymbol{\alpha}$ is obtained on classical hardware.

The rationale of the design is summarized in Table 2.

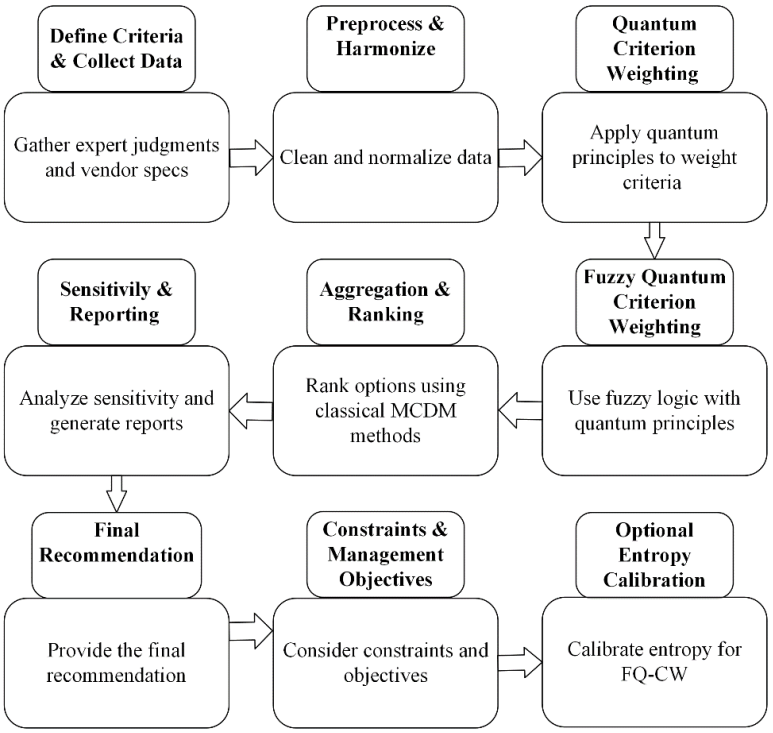


Fig. 1: Integration of QCW/FQ-CW with Classical MCDM

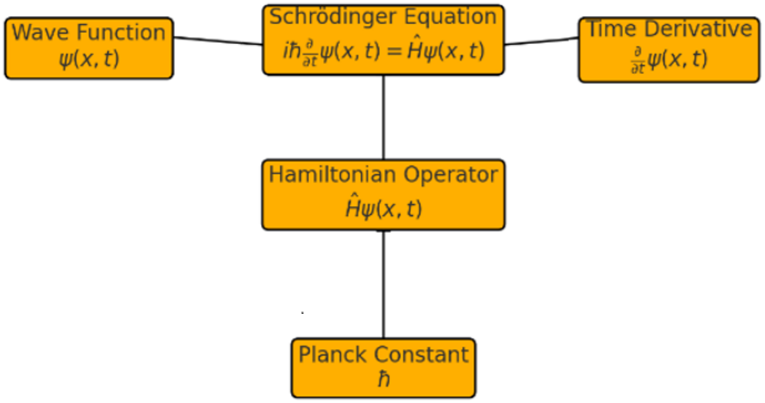


Fig. 2: Conceptual illustration of quantum state evolution (metaphor; not operationalized in the proposed algorithm)

3.5 Criterion Weighting with Fuzzy Quantum

Criterion weighting with fuzzy quantum is a novel method for multi-criteria decision-making (MCDM) problems that combines the probability-based superposition principles of quantum mechanics with the uncertainty management capabilities of fuzzy logic. While traditional methods assign deterministic weights to criteria, fuzzy quantum approaches incorporate both probabilistic and fuzzy uncertain ties simultane-

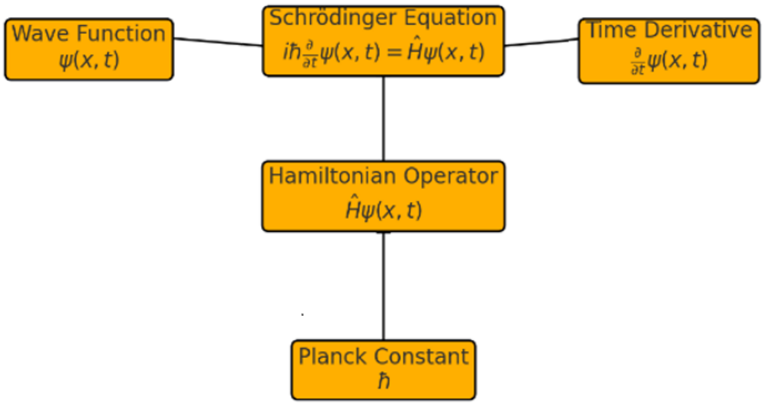


Fig. 3: Conceptual illustration of quantum state evolution (metaphor; not operationalized in the proposed algorithm)

Table 2: Rationale of Quantum-Inspired Criterion Weighting

Modeling Element	**Explanation**
Phase-angle sampling	A random phase θ_j is sampled to introduce stochasticity into the weight-generation process.
$\sin^2$ mapping	The nonnegative mapping $a_j = \sin^2(\theta_j)$ yields a probability-like mass in $[0, 1]$.
Normalization	The weight vector is normalized as $\alpha_j = a_j / \sum_k a_k$ so that $\sum_j \alpha_j = 1$.
Fuzzification (next subsection)	The probability masses are converted into triangular fuzzy weights to represent uncertainty.

ously to construct a more flexible and dynamic model. Pourabdollah et al. (2021) demonstrated that fuzzy inference systems can be considered in quantum-annealing contexts for certain optimization formulations (Pourabdollah et al., 2021). In the present study, the fuzzy quantum-inspired weighting stage is treated as a classical fuzzification of the probability masses produced by the quantum-inspired generator.

3.5.1 Quantum and Fuzzy Representation of Criteria To reflect real-world uncertainties, fuzzy triangular numbers $\tilde{c}_i = (\mu_i, \alpha_i, \beta_i)$ are used instead of classical amplitudes. In this way, the quantum decision vector is updated as follows:

$$\widetilde{\psi}(x) = \sum_i \widetilde{c}_i |K_i\rangle \tag{1}$$

The bra–ket notation is used purely as a symbolic representation of criterion indexing and does not imply a quantum-state representation in the proposed implementation.

This model is a hybrid representation that incorporates both quantum superposition and fuzzy uncertainty. The expression represents a hybrid model that includes both quantum and fuzzy variables.

3.5.2 Fuzzy Quantum-Inspired Weight Construction In the proposed framework, no unitary operators, Hamiltonians, or state-evolution equations are implemented. Instead, the quantum-inspired probability masses are transformed into triangular fuzzy numbers to represent uncertainty, and the resulting fuzzy weights are directly propagated to the MCDM stage.

3.5.3 Criterion Relationships (Scope Statement) In the present implementation, the criterion-weight generation step is performed independently for each criterion and does not include an explicit coupling mechanism. Therefore, criterion relationships are not modeled through joint amplitudes or entanglement-style constructions, and the fuzzy weights are obtained via fuzzification of the generated probability masses.

3.6 BW (Best-Worst) Weighting Method

The Best-Worst Method (BWM) is an effective multi-criteria decision-making approach that enables decision-makers to obtain more consistent and reliable weights by identifying the best and worst criteria (Ahmadi et al., 2017). The relative importance of the best criterion over the others, and how much each of the remaining criteria is better than the worst one, is expressed by the decision-maker using a scale ranging from 1 to 9. Based on these priority vectors, the following optimization model is constructed, and the criterion weights are obtained through its solution:

$$\min_{w,\xi} \xi \quad \text{s.t.} \quad \left| \frac{w_B}{w_j} - a_{Bj} \right| \leq \xi, \ \left| \frac{w_j}{w_W} - a_{jW} \right| \leq \xi \tag{2}$$

Here, w_j represents the weight of criterion j, a_Bj denotes the preference of the best criterion over criterion j, and a_{jW} denotes the preference of criterion jover the worst criterion, as formulated in Eq. (10). The BW method allows for the derivation of highly consistent weights with a minimal number of comparisons. This method is particularly preferred for large-scale problems due to its requirement for fewer pairwise comparisons.

3.7 CRITIC (Criteria Importance Through Intercriteria Correlation) Method

The CRITIC method, developed by Diakoulaki et al. (1995), offers an effective and data-driven approach for determining objective criterion weights without the need for subjective evaluation by simultaneously considering the conflict and information content among criteria (Alinezhad and Khalili, 2019). In this approach, the importance level of a criterion is calculated by combining its information content (standard deviation) and its independence from other criteria (correlation coefficient). The criterion weights are obtained by integrating the information and conflict levels for each criterion as follows:

$$C_j = \sigma_j \cdot \sum_{k=1}^{n} (1 - r_{jk}) \tag{3}$$

Here, C_j represents the information value of criterion j, σ_j is the standard deviation, and r_{jk} is the correlation coefficient between criteria j and k, as shown in Eq. (11). This method performs weight calculations entirely based on mathematical and statistical data, without requiring any subjective assessments.

3.8 Entropy Weighting Method

The entropy method is an objective technique used to determine criterion weights. This approach calculates the information content of each criterion to evaluate its importance in the decision-making process. The higher the variability and uncertainty level in the dataset, the more important the criterion becomes. On the other hand, entropy analysis methods based on the concept of information entropy introduced by Shannon allow for more accurate evaluations by measuring uncertainty and complexity, especially in cases where traditional processing methods fall short due to the noisy and incomplete structure of biomedical signals (Borowska, 2015). If a criterion exhibits high variation, it is considered to contain

more information and is therefore assigned a higher weight. In this context, the entropy value e_j and the degree of diversification d_j for each criterion are calculated, and the weights are determined as follows:

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \qquad \text{where } d_j = 1 - e_j \tag{4}$$

Here, e_j denotes the entropy of criterion j, and d_j represents the degree of diversification of that criterion, as expressed in Eq. (12). The entropy method enables objective weighting without the need for decision-makers. It is particularly powerful in large datasets for determining the importance of criteria without relying on subjective expert opinions.

3.9 MCDM (Multi-Criteria Decision-Making Methods)

Multi-Criteria Decision-Making (MCDM) is a systematic set of methods used to determine the most appropriate alternative in decision problems that involve the simultaneous evaluation of multiple criteria. In real-life decision processes, factors such as cost, efficiency, and sustainability often differ and may even conflict with each other. MCDM methods provide objective and rational evaluations by balancing these criteria in complex situations. This allows decision-makers to compare alternatives by considering both quantitative and qualitative data and to select the most suitable option. These methods are widely applied in various fields, including engineering, management, healthcare systems, and energy planning. Additionally, they offer a flexible structure through different weighting and ranking techniques.

3.9.1 CODAS (Combinative Distance-Based Assessment) Method The Combinative Distance-based Assessment (CODAS) method, developed by Keshavarz Ghorabaee et al. is employed to address multi-criteria decision-making (MCDM) problems. CODAS evaluates the distance of alternatives from the ideal solution using both Euclidean and Taxicab distances to determine their desirability. This approach provides effective and consistent results, particularly in decision-making environments that require high sensitivity (Başer and Şatır). CODAS method is one of the multi criteria decision-making (MCDM) approaches that contributes to the decision-making process by calculating the distances of alternatives from the negative ideal solution. The CODAS method ranks alternatives based on their distances to the negative ideal solution, providing the decision-maker with a clear and comparative evaluation framework (Dragana and Radojko, 2022). In this method, the distance of each alternative to the negative ideal solution is calculated using both Euclidean and Taxicab (Manhattan) distances. These two distance components are then combined to obtain the overall scores of the alternatives:

$$S_i = d_i^E + \tau d_i^T \tag{5}$$

Here, S_i represents the score of alternative i; d_i^E and d_i^T denote the Euclidean and Taxicab distances, respectively, and τ is the threshold parameter, typically chosen as 0.02, as shown in Eq. (13).

3.9.2 TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) Method The TOPSIS method is a widely used, practical, and effective decision support tool in both academic research and applied studies, as it ranks alternatives based on their closeness to the ideal solution in multi-criteria decision-making processes ([Behzadian et al., 2012). It is one of the MCDM (Multi-Criteria Decision-Making) techniques that evaluates and ranks alternatives by calculating their distances from the positive ideal and negative ideal solutions. The distance of each alternative to the positive ideal solution is denoted by d_i^+ , and its distance to the negative ideal solution is denoted by d_i^- . The TOPSIS score is then calculated using the ratio of these two distances:

$$s_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{6}$$

Here, s_i represents the TOPSIS score of alternative i, as defined ine Eq.(14). This method aims for the preferred solution to be the one that is closest to the ideal solution and farthest from the negative solution.

3.9.3 VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje) Method The VIKOR method is one of the multi-criteria decision-making (MCDM) techniques that evaluates and ranks alternatives based on their distances to the positive and negative ideal solutions. With its compromise-oriented approach, the VIKOR method offers a systematic way for decision-makers to select the alternative closest to the ideal solution by considering all types of criteria. It is particularly effective in fields that require precision, such as engineering and biomedical applications (Jahan et al., 2011). In this method, the total benefit difference (S_i) and the maximum criterion difference (R_i) are calculated for each alternative, and a compromise solution score Q_i is obtained based on these values. The VIKOR scores are computed using the following formula:

$$Q_i = v\frac{S_i - S_{\min}}{S_{\max} - S_{\min}} + (1-v)\frac{R_i - R_{\min}}{R_{\max} - R_{\min}} \tag{7}$$

Here, v is the compromise coefficient, typically set to 0.5. The VIKOR method provides a balanced solution approach for multi-criteria problems, focusing on group benefit rather than individual optimum.

3.9.4 MOORA (Multi-Objective Optimization on the Basis of Ratio Analysis) Method The MOORA method is one of the multi-criteria decision-making (MCDM) techniques that ranks decision alternatives based on the weighted normalized values of criteria. By employing a multi objective decision-making structure based on ratio analysis, the MOORA method provides practical, flexible, and effective solutions to complex selection problems encountered in production environments, demonstrating high applicability in real-time decision-making processes (Chakraborty, 2011). In this method, benefit and cost criteria are evaluated separately, and their combined effect is calculated to obtain a net score. The MOORA score is calculated as follows:

$$MOORA_i = S_i^+ - S_i^- \tag{8}$$

Here, S_i^+ represents the total value of benefit criteria, while S_i^- represents the total value of cost criteria, as defined in Eq.(16).

3.9.5 Fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) Method The Fuzzy TOPSIS method is developed to better represent decision-makers' judgments in uncertain decision environments. It preserves the advantages of classical TOPSIS while offering more flexible and interpretable solutions to multi-criteria decision-making problems through fuzzy sets (Nădăban et al., 2016). In this method, criterion values are expressed as fuzzy numbers (typically triangular fuzzy numbers TFNs), and the distances of alternatives from the fuzzy positive and fuzzy negative ideal solutions are calculated. Alternatives are then ranked based on a similarity coefficient. The Fuzzy TOPSIS score is calculated as follows:

$$CC_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{9}$$

Here, CC_i denotes the fuzzy similarity coefficient of alternative i, while d_i^+ and d_i^- represent the distances to the fuzzy positive and fuzzy negative ideal solutions, respectively, as shown in Eq.(17).

3.9.6 Fuzzy EDAS (Evaluation based on Distance from Average Solution) Method The Fuzzy EDAS method is one of the multi-criteria decision-making (MCDM) techniques that evaluates and ranks alternatives based on their distances to the positive and negative ideal solutions. Fuzzy EDAS provides both flexibility and accuracy in multi-criteria group decision-making processes characterized by conflicting attributes and uncertainty, by evaluating alternatives based on their distance from the average solution (Jana and Pal, 2021). In this method, criterion values are represented using triangular fuzzy numbers (TFNs), and for each alternative, positive distance (PD) and negative distance (ND) scores are calculated. These values are then multiplied by their respective weights and normalized to obtain the final EDAS score:

$$S_i = \frac{1}{2}\left(\frac{PDS_i}{\max PDS_i} + \frac{NDS_i}{\max NDS_i}\right) \tag{10}$$

Here, PDS_i and NDS_i represent the positive and negative distance scores of alternative i, respectively, and S_i denotes the final fuzzy EDAS score, as defined in Eq.(18).

3.9.7 Fuzzy VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje) Method The Fuzzy VIKOR method is one of the multi-criteria decision-making (MCDM) techniques that evaluates and ranks alternatives based on their distances to the positive and negative ideal solutions. The Fuzzy VIKOR method provides a rational and systematic decision-making approach that aims to comprehensively rank the performance of feasible alternatives by considering the weights of evaluation criteria in multi-criteria decision environments characterized by uncertainty (Chang, 2014). In this method, the criteria are expressed using triangular fuzzy numbers (TFNs), and for each alternative, the total benefit difference (S_i) and the maximum criterion difference (R_i) are calculated with respect to the positive and negative ideal solutions. Based on these two values, the Q_i score—representing the fuzzy compromise solution—is obtained using the formulation presented in Eq. (19). This equation integrates both benefit and regret measures through the compromise coefficient v, which is typically set to 0.5, to provide a balanced evaluation among competing alternatives:

$$Q_i = v\frac{S_i - S_{\min}}{S_{\max} - S_{\min}} + (1-v)\frac{R_i - R_{\min}}{R_{\max} - R_{\min}} \tag{11}$$

4 Conceptual Discussion: Quantum–Fuzzy Decision-Making (Non-operational)

This section provides conceptual interpretation only and does not describe implemented algorithmic mechanisms. The proposed method is a classical, quantum-inspired stochastic fuzzy weighting approach; therefore, the discussion below is included solely to contextualize common terminology used in the related literature.

Figure 4 is included as a conceptual narrative frequently used in quantum-inspired decision-model discussions (e.g., superposition/interference/collapse). It is not intended to represent an implemented sequence of algorithmic steps in the proposed method.

Figure 5 is presented as a high-level conceptual map and should not be interpreted as evidence of improved stability, discrimination, or performance of the proposed method in the absence of a formal robustness analysis or benchmarking.

4.1 Weighting Models and Decision Outcomes (Conceptual, Non-operational)

Quantum terminology (e.g., superposition or entanglement) is often used in the literature to motivate alternative probabilistic viewpoints; however, the present implementation does not simulate quantum states and does not include explicit coupling between criteria.

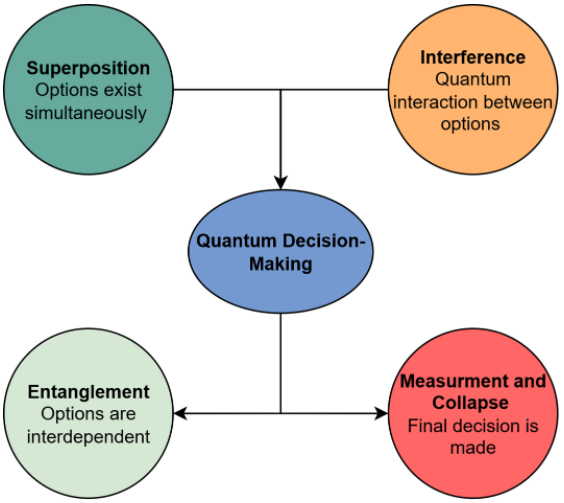


Fig. 4: Conceptual illustration of a quantum decision-making narrative (non-operational)

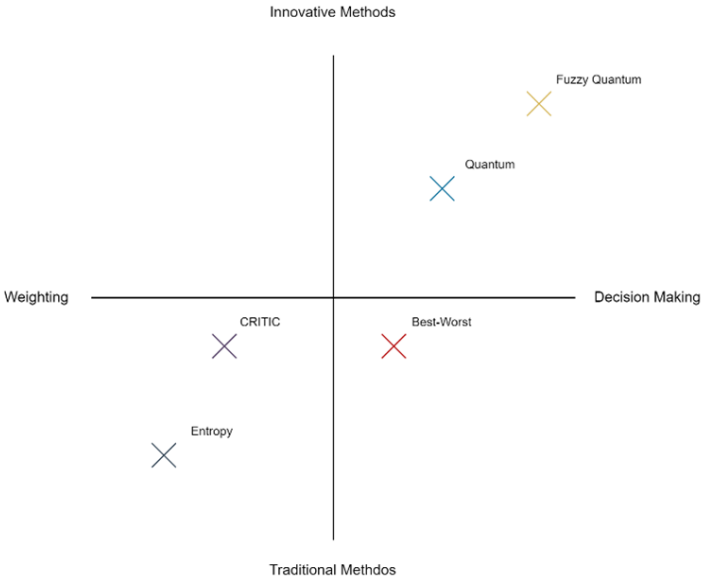


Fig. 5: Conceptual positioning of related ideas (non-operational; not an implemented framework)

In the implemented study, differences in alternative rankings are primarily driven by the selected weighting scheme (BW, CRITIC, Entropy, quantum-inspired, and fuzzy quantum-inspired) and by the downstream MCDM method; therefore, claims regarding stability, discrimination, or interdependency modeling are not asserted as formal properties without dedicated robustness metrics and benchmarking.

4.2 Algorithm: Quantum–Fuzzy Weighting Integration Step

The integration of quantum and fuzzy logic within the weighting process provides a robust framework for managing uncertainty and interdependence among criteria in multi-criteria decision-making (MCDM)

systems. By combining the probabilistic representation of quantum mechanics with the linguistic flexibility of fuzzy logic, this algorithm enables a dynamic and adaptive weighting structure. The following steps formalize the process through which normalized decision data are enhanced with quantum-derived probability amplitudes and transformed into triangular fuzzy weights, ensuring that both the uncertainty and interaction effects are preserved before being transferred to the MCDM evaluation stage.

Step 1: Criterion direction and scale alignment (normalization)

$$r_{ij} = \begin{cases} \dfrac{x_{ij}}{\max_i x_{ij}}, C_j(\text{benefit}) \\ \dfrac{\min_i x_{ij}}{x_{ij}}, C_j(\text{cost}) \end{cases} \tag{12}$$

Equation (20) shows the normalization step that aligns all criteria to a common scale and direction. Benefit criteria are divided by the column maximum, while cost criteria use the column minimum. This ensures all criteria are comparable within the [0,1] range.

Step 2: Quantum-originated probability mass generation

$$\theta_j \sim U(0,\pi), \qquad a_j = \sin^2(\theta_j), \qquad \alpha_j = \frac{a_j}{\sum_{k=1}^{n} a_k}, \qquad \sum_j \alpha_j = 1 \tag{13}$$

Equation (21) generates a quantum-based probability distribution for each criterion. θ represents the random phase angle, while α denotes the normalized probability weight derived from the superposition effect.

Step 3: Fuzzification of quantum probability (TFN weighting)

$$\tilde{w}_j = \big(\max(\alpha_j - \delta, 0),\ \alpha_j,\ \min(\alpha_j + \delta, 1)\big) \tag{14}$$

Equation (22) represents the transformation of quantum probability weights into triangular fuzzy numbers (TFNs). α_j is the central value, and δ defines the uncertainty range, allowing the probability distribution to capture human judgment ambiguity.

Step 4: Normalized decision matrix with TFN

$$\tilde{r}_{ij} = r_{ij} \otimes \tilde{w}_{ij} = \big(r_{ij}\, l_j^{(w)},\ r_{ij}\, m_j^{(w)},\ r_{ij}\, u_j^{(w)}\big) \tag{15}$$

Equation (23) shows the multiplication of normalized criterion values with quantum–fuzzy weights to form the weighted decision matrix. As a result, each element is represented as a triangular fuzzy number (TFN).

Step 5: Reference Point Generation for MCDM Methods

$$\tilde{v}_j^{(k)} = \Phi_j^{(k)}\big(\{\tilde{r}_{ij}\}_{i=1}^{m}; \theta_j\big), \qquad k = 1, 2, \ldots, K_j \tag{16}$$

In Equation (24), for each criterion, the method-specific reference point(s) are generated from the set of normalized–weighted values $(\tilde{r}_{ij})$. This process is expressed by a general generation function ϕ, defined according to method- or criterion-specific parameters θ_j. The resulting $\tilde{v}_j^{(k)}$ values form the fundamental reference points used by MCDM methods.

Step 6: Distance / difference measures between TFNs

$$d_E\big((l_1, m_1, u_1), (l_2, m_2, u_2)\big) = \sqrt{(l_1 - l_2)^2 + (m_1 - m_2)^2 + (u_1 - u_2)^2} \tag{17}$$

$$d_T\big((l_1, m_1, u_1), (l_2, m_2, u_2)\big) = |l_1 - l_2| + |m_1 - m_2| + |u_1 - u_2| \tag{18}$$

Equation (25) and (26) represents the distance between two triangular fuzzy numbers (TFNs). d_E denotes the fuzzy Euclidean distance, while d_T refers to the fuzzy Taxicab (Manhattan) distance. These measures are used to determine how far each alternative is from the reference point.

Step 7: Fuzzy Quantum Weighted Decision Matrix and MCDM Integration

$$S_i = f(\tilde{R}, \tilde{W}) = f(r_{ij} \times \tilde{w}_j) \tag{19}$$

In Equation (27), this expression represents the integration of normalized criterion values (r_ij) with quantum–fuzzy weights $\tilde{w}_j$and their transfer into the multi-criteria decision-making (MCDM) model. The function f(.)varies depending on the selected method (e.g., CODAS, EDAS, TOPSIS, VIKOR, etc.) and is used to calculate the final scores (S_i) of the alternatives. At this stage, the resulting quantum–fuzzy weighted decision matrix forms a common decision space utilized for evaluation across all MCDM approaches.

Step 8: Ranking alternatives

$$A_{(1)} = \arg\max_i S_i, \qquad A_{(2)}, A_{(3)}, \ldots, A_{(m)} \tag{20}$$

In Equation (28) represents the ranking of alternatives in multi-criteria decision-making (MCDM) methods based on their final scores (S_i). All alternatives are ordered from the highest to the lowest score, and the one with the highest value $A_{(1)}$is selected as the most suitable alternative. This ranking approach is universally applicable to all MCDM methods.

5 Result

In this study, a comprehensive analysis was conducted using different weighting methods and Multi Criteria Decision-Making (MCDM) techniques to address the industrial robot selection problem. Both classical MCDM methods such as CODAS, TOPSIS, VIKOR, and MOORA, and fuzzy logic-based MCDM methods (fuzzy TOPSIS, fuzzy EDAS, fuzzy VIKOR) were applied to evaluate alternative robots and determine the most suitable selection. For each method, BW, CRITIC, Entropy, Quantum, and Fuzzy Quantum weighting techniques were employed, and the obtained results were analyzed comparatively.

Table 3: Best Robot Selection with Different Weighting Methods (Fuzzy MCDM Methods)

Weighting Method	**Fuzzy TOPSIS**		**Fuzzy EDAS**		**Fuzzy VIKOR**	
	Robot	Score	Robot	Score	Robot	Score
BW	IRB 5720-180/2.6	0.539944	IRB 760	5.385003	IRB 140-6/0.8	0.000000
CRITIC	IRB 1010-1.5/0.37	0.600000	IRB 760	2.135629	IRB 140-6/0.8	0.024323
Entropy	IRB 760	0.524325	IRB 760	5.589297	IRB 1600-6/1.2	0.010121
Quantum	IRB 1010-1.5/0.37	0.556295	IRB 1010-1.5/0.37	2.679314	IRB 140-6/0.8	0.066591
Fuzzy Quantum	IRB 1010-1.5/0.37	0.556331	IRB 1010-1.5/0.37	2.682965	IRB 140-6/0.8	0.049457

The results in Table 3 show that the optimal robot selection varies depending on the fuzzy MCDM method and weighting technique applied. While Fuzzy TOPSIS and Fuzzy Quantum methods highlighted the IRB 1010-1.5/0.37 model, Fuzzy EDAS selected IRB 760, and Fuzzy VIKOR identified IRB 140-6/0.8 as the best alternative. Moreover, the quantum-inspired and fuzzy quantum-inspired weighting schemes yielded rankings that differed from those produced by Entropy and CRITIC, indicating that the ranking outcome is sensitive to the adopted weighting design.

According to Table 4, while classical methods highlighted the IRB 760 model, the VIKOR method identified IRB 140-6/0.8 as the best alternative. However, when Quantum and Fuzzy Quantum weighting techniques were applied, the IRB 1010-1.5/0.37 model became more advantageous in several methods. These findings indicate that the selected alternative may change as a function of the adopted weighting scheme.

Table 4: Best Robot Selection with Different Weighting Methods (Classical MCDM Methods)

Weighting Method	**CODAS**		**TOPSIS**		**VIKOR**		**MOORA**	
	Robot	Score	Robot	Score	Robot	Score	Robot	Score
BW	IRB 760	0.492034	IRB 760	0.703721	IRB 140-6/0.8	0.000000	IRB 760	0.276357
CRITIC	IRB 760	0.345672	IRB 760	0.549528	IRB 140-6/0.8	0.058810	IRB 760	0.069460
Entropy	IRB 760	0.469126	IRB 760	0.554542	IRB 2000/1	0.029126	IRB 760	0.209345
Quantum	IRB 1010-1.5/0.37	0.379092	IRB 1010-1.5/0.37	0.556295	IRB 140-6/0.8	0.108056	IRB 5400-22	0.503426
Fuzzy Quantum	IRB 1010-1.5/0.37	0.381487	IRB 1010-1.5/0.37	0.556331	IRB 140-6/0.8	0.049457	IRB 5400-22	0.503426

5.1 Comparison of Classical, Quantum and Fuzzy Quantum Weighting

Weighting techniques play a critical role in the effectiveness of multi-criteria decision-making (MCDM) models. In recent years, advanced methods such as quantum and fuzzy quantum-based approaches have gained attention for their ability to model uncertainty and complex interactions. This section evaluates the impact of various weighting methods on decision quality through a comparative analysis.

Table 5: Comparison of Quantum, Fuzzy Quantum and Classical Weighting Methods

Feature	**Classical Weighting**	**Quantum Weighting**	**Fuzzy Quantum Weighting (Proposed)**
Uncertainty Handling	Fixed weights (deterministic)	Probability-based weights via superposition principle	Both quantum probabilities and fuzzy uncertainties are considered
Relationships Between Criteria	Independent criteria	Independent criteria (no coupling implemented)	Independent criteria with fuzzy uncertainty modeling
Weight Updating	Fixed or classically updated	Stochastically generated by phase-angle sampling	Stochastically generated and fuzzified as TFNs
Stability of Results	Singular and fixed result	Probabilistic weights within repeated runs	Probabilistic and fuzzy weights within repeated runs
Computational Complexity	Low	Low–medium (classical sampling)	Low–medium (sampling + fuzzification)
Dynamic Decision-Making	None (static system)	Provides stochastic variability via repeated sampling	Provides stochastic and fuzzy variability (sampling + TFN fuzzification)
Real-World Applications	Suitable for fixed decision process	Provides an alternative modeling perspective for uncertainty	Provides an alternative modeling perspective with fuzzy uncertainty representation

As shown in Table 5, the weighting design plays a crucial role in MCDM outcomes, making the comparative analysis of weighting methods particularly important. The table presents a comparison of classical, quantum-inspired, and fuzzy quantum-inspired weighting approaches across several dimensions. Classical weighting techniques rely on fixed and deterministic values and assume independence between criteria, making them suitable for static decision environments. In contrast, the quantum-inspired approach generates probabilistic weights through classical sampling and normalization, whereas the fuzzy quantum-inspired approach additionally represents uncertainty by fuzzifying these weights as TFNs.

6 Discussion

In this study, the industrial robot selection problem has been examined under multiple classical and fuzzy MCDM procedures while systematically varying the criterion-weighting scheme. The quantum-inspired and fuzzy quantum-inspired schemes have functioned as stochastic generators of probability masses and their fuzzy counterparts, and the resulting weights have been propagated through standard ranking mechanisms.

The comparative results indicate that the identified best alternative is sensitive to the adopted weighting method, which is consistent with the well-known role of weighting in MCDM. In particular, the quantum-inspired schemes have produced rankings that differ from those obtained by Entropy and CRITIC, suggesting that stochastic probability-mass generation can act as an alternative regularization mechanism in the weighting stage. Moreover, the fuzzy extension has been observed to support uncertainty representation by assigning TFN-valued weights rather than point estimates.

It is emphasized that the proposed approach has been implemented entirely on classical hardware and has not operationalized quantum-mechanical coupling mechanisms between criteria (e.g., entanglement-based joint amplitudes, unitary evolution, or Schrödinger dynamics). Therefore, the contribution has been restricted to a classical, quantum-inspired probabilistic fuzzy weighting design and its empirical integration with established MCDM methods.

With respect to the research questions, the effect of the quantum-inspired weighting on decision outcomes (RQ1) has been assessed via cross-method ranking comparisons; the role of fuzzification under uncertainty (RQ2) has been examined by TFN-based weighting and fuzzy MCDM outputs; and ranking consistency patterns associated with the fuzzy quantum-inspired scheme (RQ3) have been discussed by comparing best-alternative consistency across methods.

It is noted that these consistency patterns are discussed qualitatively and are not presented as formal statistical robustness properties in the absence of variance estimates, confidence intervals, or dedicated stability metrics.

A key limitation is that no computational complexity analysis or runtime benchmarking has been provided, and thus no performance claims beyond methodological modeling have been made. Future work may include (i) formal sensitivity/stability analyses under repeated sampling, (ii) explicit modeling of inter-criterion dependence through correlated sampling or coupling terms, and (iii) runtime evaluations on larger datasets to characterize practical scalability.

7 Conclusion and Future Works

In this study, a classical, quantum-inspired probabilistic weighting mechanism has been integrated with fuzzy uncertainty modeling and combined with conventional MCDM ranking methods for industrial robot selection. The weighting stage has been constructed by phase-angle sampling followed by $\sin^2(\cdot)$ mapping and normalization, and the resulting probability masses have been fuzzified as triangular fuzzy numbers when the fuzzy quantum-inspired variant has been applied.

It has been observed that the selected best alternative and the stability of rankings may vary across weighting schemes and MCDM procedures. In this respect, the proposed quantum-inspired and fuzzy quantum-inspired schemes have been shown to provide an additional perspective for sensitivity analysis in criterion weighting under uncertainty, without requiring subjective expert elicitation.

It is emphasized that inter-criterion dependence has not been explicitly modeled and that the presented framework does not implement quantum operators, unitary evolution, entanglement-based coupling, qubit simulation, or hardware-based quantum computation. Accordingly, the scope of the contribution has been restricted to classical decision modeling.

For future work, (i) correlated or copula-based sampling strategies may be introduced to represent inter-criterion dependence, (ii) convergence and stability under repeated sampling may be evaluated formally, and (iii) computational cost and scalability may be assessed through runtime benchmarking on larger and more diverse datasets.

Acknowledgment

The author would like to thank the reviewers for their insightful comments and suggestions, which significantly improved the quality of this paper.

References

Abdelgaber, N., Nikolopoulos, C. (2020, December). Overview on quantum computing and its applications in artificial intelligence. In *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)* (pp. 198–199). IEEE.

Ahmadi, H. B., Kusi-Sarpong, S., Rezaei, J. (2017). Assessing the social sustainability of supply chains using Best Worst Method. *Resources, Conservation and Recycling*, 126, 99–106.

Alinezhad, A., Khalili, J. (2019). New methods and applications in multiple attribute decision making (MADM) (Vol. 277, pp. 199–203). Cham: Springer.

Arslankaya, D., Göraltay, K. (2019). Çok kriterli karar verme yöntemlerinde güncel yaklaşımlar.

Başer, E., Şatır, Ö. Ü. E. Kuantum Hesaplama.

Behzadian, M., Otaghsara, S. K., Yazdani, M., Ignatius, J. (2012). A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39(17), 13051–13069.

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671), 195–202.

Borowska, M. (2015). Entropy-based algorithms in the analysis of biomedical signals. *Studies in Logic, Grammar and Rhetoric*, 43(1), 21–32.

Busemeyer, J. R., Wang, Z., Townsend, J. T. (2006). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, 50(3), 220–241.

Chakraborty, S., Datta, H. N., Kalita, K., Chakraborty, S. (2023). A narrative review of MOORA in decision making. *Opsearch*, 60(4), 1844–1887.

Chakraborty, S. (2011). Applications of the MOORA method for decision making in manufacturing environment. *IJAMT*, 54(9), 1155–1166.

Chang, T. H. (2014). Fuzzy VIKOR method: A case study of the hospital service evaluation in Taiwan. *Information Sciences*, 271, 196–212.

Chella, A., Gaglio, S., Pilato, G., Vella, F., Zammuto, S. (2022). A quantum planner for robot motion. *Mathematics*, 10(14), 2475.

Chodha, V., Dubey, R., Kumar, R., Singh, S., Kaur, S. (2022). Selection of industrial arc welding robot with TOPSIS and Entropy MCDM techniques. *Materials Today: Proceedings*, 50, 709–715.

Chu, T. C., Lin, Y. C. (2003). A fuzzy TOPSIS method for robot selection. *IJAMT*, 21(4), 284–290.

Coccia, M. (2024). Converging AI and quantum technologies. *Technologies*, 12(5), 66.

Çolak, M., Kaya, İ. (2017). Prioritization of renewable energy alternatives by an integrated fuzzy MCDM model. *Renewable and Sustainable Energy Reviews*, 80, 840–853.

Dasgupta, S. (2024). Stability of quantum computers. *arXiv:2404.19082*.

Değirmenci, M., Utku, Ş. (2000). Yönetim ve örgüt yapısına kuantum mekaniği açısından bir bakış. *Doğuş Üniversitesi Dergisi*, 1(2), 76–83.

Dragana, V. K., Radojko, L. U. K. I. Ć. (2022). Efficiency analysis of agriculture in Serbia based on the CODAS method. *International Review*, (1–2), 33.

Du, J., Li, H., Xu, X., Shi, M., Wu, J., Zhou, X., Han, R. (2003). Experimental implementation of the quantum random-walk algorithm. *Physical Review A*, 67(4), 042316.

Dunjko, V., Briegel, H. J. (2018). Machine learning & AI in the quantum domain. *Reports on Progress in Physics*, 81(7), 074001.

Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K., Lukens, J. E. (2000). Quantum superposition of distinct macroscopic states. *Nature*, 406(6791), 43–46.

Ghosh, S., Opala, A., Matuszewski, M., Paterek, T., Liew, T. C. (2019). Quantum reservoir processing. *npj Quantum Information*, 5(1), 35.

Goh, C. H. (1997). Analytic hierarchy process for robot selection. *Journal of Manufacturing Systems*, 16(5), 381–386.

Hayal Gerçeklerinin Gücü. Kuantum Bilgisayarlar.

Hadap, S., Patil, M. (2024). Quantum Computing in Artificial Intelligence: A Paradigm Shift. *IJARSCT*, 530–534.

Harrow, A. W., Hassidim, A., Lloyd, S. (2009). Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15), 150502.

Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K. (2009). Quantum entanglement. *Reviews of Modern Physics*, 81(2), 865–942.

Jahan, A., Mustapha, F., Ismail, M. Y., Sapuan, S. M., Bahraminasab, M. (2011). A comprehensive VIKOR method for material selection. *Materials & Design*, 32(3), 1215–1221.

Jain, Y. K., Likhitkar, B., Verma, P., Mishra, P. K., Rathore, D. S., Bhokare, M. S. (2025, March). Leveraging Quantum Computing for MCDM in Strategic Business Planning. In *AUTOCOM 2025* (pp. 1492–1496). IEEE.

Jana, C., Pal, M. (2021). Extended bipolar fuzzy EDAS for group decision-making. *Computational and Applied Mathematics*, 40(1), 9.

Kağızman, A., Deveci, K., Sezer, V. (2022). Selection of Suitable Automatic CPR Device Chassis Material Using Intuitionistic Fuzzy TOPSIS, VIKOR, and CODAS Methods.

Karaatlı, M., & Dağ, O. (2018). Türk Milli Erkek Futbol Takımına ÇKKV ile futbolcu seçimi. *SDU IIBF Dergisi*, 23(4), 1433–1454.

Koçhan, M. (2024). Kuantum fiziğinin bulanık mantık perspektifinden tasviri. *Hitit İlahiyat Dergisi*, 23(1), 300–335.

Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., Bansal, R. C. (2017). A review of MCDM towards sustainable renewable energy. *Renewable and Sustainable Energy Reviews*, 69, 596–609.

Lamata, L. (2017). Basic protocols in quantum reinforcement learning with superconducting circuits. *Scientific Reports*, 7(1), 1609.

Lohia, A. (2024). Quantum AI: Enhancing ML with Quantum Computing. *Journal of Quantum Science and Technology*, 1(2), 6–11.

Merzbacher, E. (1998). *Quantum Mechanics*. John Wiley & Sons.

Nădăban, S., Dzitac, S., Dzitac, I. (2016). Fuzzy TOPSIS: a general view. *Procedia Computer Science*, 91, 823–831.

Nagaraj, G., Upadhayaya, N., Matroud, A., Sabitha, N., VK, R., Nagaraju, R. (2023, March). Impact of quantum computing on improving AI. In *ICIDCA 2023* (pp. 447–452). IEEE.

Pourabdollah, A., Acampora, G., Schiattarella, R. (2021). Fuzzy logic on quantum annealers. *IEEE Transactions on Fuzzy Systems*, 30(8), 3389–3394.

Pyrkov, A., Aliper, A., Bezrukov, D., Lin, Y. C., Polykovskiy, D., Kamya, P., Zhavoronkov, A. (2023). Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discovery Today*, 28(8), 103675.

Schuld, M., Sinayskiy, I., Petruccione, F. (2015). An introduction to quantum machine learning. *Contemporary Physics*, 56(2), 172–185.

Singh, P. (2021). FQTSFM: A fuzzy-quantum time series forecasting model. *Information Sciences*, 566, 57–79.

Tiwari, P., Zhang, L., Qu, Z., Muhammad, G. (2024). Quantum fuzzy neural network for multimodal sentiment and sarcasm detection. *Information Fusion*, 103, 102085.

Vadisetty, R., Polamarasetti, A. (2024, November). Quantum Computing For Cryptographic Security With AI. In *ICCMA 2024* (pp. 252–260). IEEE.

Wankhede, V. A., Vinodh, S. (2021). Analysis of Industry 4.0 challenges using best worst method: A case study. *Computers & Industrial Engineering*, 159, 107487.

Wei, C., Wu, J., Guo, Y., Wei, G. (2021). Green supplier selection based on CODAS in probabilistic uncertain linguistic environment. *Technological and Economic Development of Economy*, 27(3), 530–549.

Wichert, A. (2020). *Principles of Quantum Artificial Intelligence: Quantum Problem Solving and Machine Learning*.

Wiebe, N., Kapoor, A., Svore, K. M. (2014). Quantum deep learning. *arXiv:1412.3489*.

Xiong, Q., Fei, Y., Du, Q., Zhao, B., Di, S., Shan, Z. (2024). A modified lightweight quantum CNN for malicious code detection. *Quantum Science and Technology*, 10(1), 015007.

Yang, X., Chen, Z. (2024). Exploiting temporal features in water resource carrying capacity assessment based on extended VIKOR. *Sustainable Cities and Society*, 114, 105801.

Zheng, P., Zubatyuk, R., Wu, W., Isayev, O., Dral, P. O. (2021). AI-enhanced quantum chemical method with broad applicability. *Nature Communications*, 12(1), 7022.

Zhong, S., Chen, Y., Miao, Y. (2023). Using improved CRITIC method to evaluate thermal coal suppliers. *Scientific Reports*, 13(1), 195.

Received July 28, 2025 , revised February 1, 2026, accepted February 10, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, pp. 174–189
https://doi.org/10.22364/bjmc.2026.14.1.09

Exploring Hybrid Quantum-Classical Methods for Practical Time-Series Forecasting

Maksims DIMITRIJEVS, Mārtiņš KĀLIS, Iļja REPKO

Center for Quantum Computer Science, Faculty of Sciences and Technology, University of Latvia *

maksims.dimitrijevs@lu.lv, martins@kalis.lv, ilja.repko@gmail.com

ORCID 0000-0002-4225-7889, ORCID 0000-0001-9426-9497, ORCID 0009-0003-3154-1803

Abstract. Time-series forecasting is essential for strategic planning and resource allocation, and is important in areas like finance, supply chain management, and energy demand prediction. In such areas, small improvements in forecast quality can lead to significant economic and strategic benefits. It is important to assess the practicality of quantum-enhanced forecasting methods because quantum approaches can complement the strengths of classical models.
In this work, we explore two quantum-based approaches for time-series forecasting. The first approach utilizes a parameterized quantum circuit (PQC) model. The second approach employs variational quantum linear regression (VQLS), formulating the forecasting problem as a system of linear equations solved via quantum optimization. A comparative analysis is conducted to assess their practical applicability and performance.
In our experiments, the PQC with COBYLA optimization achieved the lowest test error rates, outperforming the classical baselines under the reported settings. VQLS formulation faced stability issues on general synthetic datasets.

Keywords: quantum, time-series, PQC, VQLS, forecasting

1 Introduction

Time-series forecasting is a critical tool in fields where accurate predictions of future values based on past observations can drive strategic decision-making and resource optimization (Fisher et al., 2024), (Masini et al., 2020). The fundamental objective in time-series forecasting is to develop a model that effectively captures underlying patterns in historical data and extrapolates them to future points (Fisher et al., 2024). One classical approach to this problem involves framing the prediction as a regression task, where a sliding window of previous time-steps is used to forecast the next value in

* Project conducted in collaboration with Forse AI (Sapiens.BI SIA)

the series (Masini et al., 2020), (Lara-Benítez et al., 2021). This approach naturally leads to a linear system, where past values in the time series are combined through a set of coefficients to produce the forecast. Solving this system by determining the optimal coefficients allows for efficient predictions, and thus, the time-series forecasting task becomes closely aligned with finding solutions to linear equations – a process that is computationally intensive for large datasets but offers promising advantages when adapted to quantum computing techniques (Harrow et al., 2009).

Quantum computing harnesses the principles of quantum mechanics to process information. Unlike classical computers that use bits (0 or 1), quantum computers utilize qubits. Qubits can exist in a superposition, representing both 0 and 1 simultaneously. This fundamental difference allows quantum computers to perform computations in parallel, potentially leading to significant speedups for certain problem classes. However, one has to measure the quantum state to learn any results, and the *interesting* results are only measured with some probability. The goal of quantum algorithms is to increase the probability that the *interesting* results are measured with high enough probability.

While the potential of quantum computing is considerable, current technology faces limitations. Currently, we are in the era of Noisy Intermediate-Scale Quantum (NISQ) devices (Preskill, 2018). These devices are susceptible to noise and have limited qubit counts, hindering large-scale calculations. However, even with these constraints, researchers are exploring algorithms and applications suitable for NISQ devices, such as hybrid quantum-classical approaches.

Algorithms such as the Harrow–Hassidim–Lloyd (HHL) algorithm offer an exponential speed-up over the best known classical algorithms, but they require large and fault-tolerant quantum computers that are not available currently (Harrow et al., 2009). Hybrid quantum-classical algorithms like the variational quantum linear solver (VQLS) (Bravo-Prieto et al., 2023) aim to mitigate the limitations of NISQ devices. These algorithms combine short-depth quantum circuits for cost function evaluation with classical optimization techniques, reducing the reliance on extensive quantum resources.

The performance of both the HHL and VQLS algorithms depends on the condition number of the coefficient matrix of the linear system. There are no guarantees for the coefficient matrices of time-series problems to be well-conditioned. In fact, they are likely to have high condition numbers, leading to weak performance guarantees. However, many models in machine learning do not have strong theoretical performance guarantees, but perform well in practice nonetheless (Zhang et. al., 2021), (Halevy et. al., 2009). In this paper, we set out to test the practical performance of a quantum version of a linear system solver.

Summary of the main contributions of our paper:

- We implemented and evaluated two hybrid quantum-classical approaches for time-series forecasting: a parameterized quantum circuit (PQC) regressor and a VQLS-based regression model.
- We compared quantum models against the classical baselines (linear model, ridge regression, and small neural network) under similar experimental conditions.

- We provided a detailed analysis of model performance and test error for different metrics. This highlights the cases where quantum methods are competitive and where they face limitations.
- We discussed practical considerations such as optimization, reproducibility, and scalability. Therefore, we situate our findings within the broader context of quantum machine learning for time-series forecasting.

The remainder of the paper is organized as follows. Section 2 introduces the methods, including the PQC, the classical regression baselines, and the VQLS formulations. Section 3 presents the experimental setup and evaluation metrics, followed by the results and comparative analysis. Section 4 is the conclusion section that includes a comprehensive summary and a discussion of the findings.

The codes and results of our experiments are publicly available: `https://github.com/infenrio/pqc_and_vqls`.

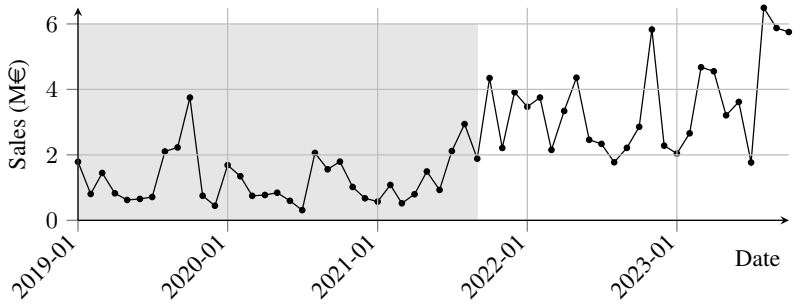
2 Methods

In this section, we introduce the methodological framework of our study. We describe the quantum models (PQC and VQLS) and the classical baselines, along with the data preprocessing steps, problem formulation, and some technical details of the implementation.

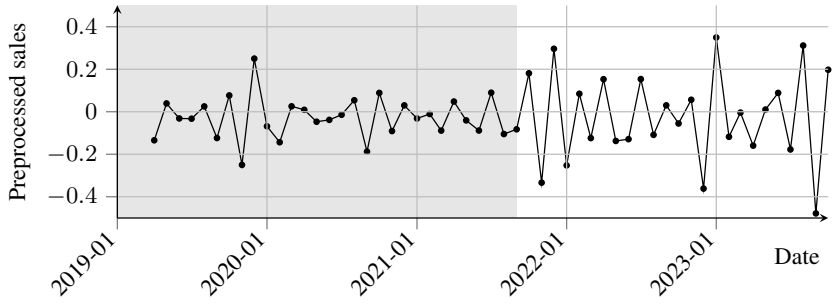
The study evaluated the performance of a PQC model (Benedetti et al., 2019), where two optimization algorithms, L-BFGS-B and COBYLA, were employed to find optimal parameters. The PQC's performance was compared against two classical baselines: a linear regression model and a simple deep learning model. The deep learning model consisted of two hidden layers, each with 12 units and ReLU activation, to mirror the limitations of the currently available quantum devices. The primary evaluation metric used across all models was the mean squared error (MSE).

The dataset used in the experiments was provided by *Forse AI*. It is a synthetic dataset designed to mimic typical sales data patterns analyzed by the company.

Standard data preprocessing steps were applied to ensure the data was suitable for both classical and quantum models. To address non-stationarity often present in time-series sales data, differencing was performed, which enhanced the stationarity of the dataset. This step is critical for improving the performance of models that assume stationarity in their inputs. Furthermore, the data was scaled to fall within a normalized range, a necessary condition to facilitate encoding on a quantum device, which operates effectively only on appropriately scaled data. This preprocessing ensures compatibility and optimal performance across all evaluated models.



(a) Original data. Data in the shaded area is used for training and scaling adjustments. The rest is used for testing.



(b) Preprocessed data after applying differencing (change in sales vs. absolute sales) and scaling the differences to the range [-0.25, +0.25]. Note that test-time data can fall outside of this range.

Figure 1. Comparison of monthly sales (M€) before and after preprocessing from 2019 to 2023.

2.1 Parameterized quantum circuit

The core of the algorithm lies in the PQC, which consists of two primary components:

- *Feature map* encodes the features of each training sample as angles of rotation gates applied to the qubits. In the experiments, 12 previous time-steps are used as features, encompassing information on a full year.
- *Parameterized ansatz* is comprised of entangling gates (CNOT in this case) and parameterized single-qubit gates (R_X and R_Y). The parameters within these gates are what the algorithm adjusts during training to align the predicted values with the actual values from the training set.

The algorithm learns by iteratively updating the parameters of the ansatz to minimize the squared difference between the expected value of an observable (calculated using a Qiskit simulator) and the actual labels for each training sample.

PQC consists of 12 qubits, feature map encodes features through 12 rotations, ansatz has two layers, each consisting of one layer of two-qubit CNOT gates, and two different rotations on each qubit (see Figure 2). Ansatz has a circuit depth of 6 layers of gates and 48 angles that serve as learnable parameters.

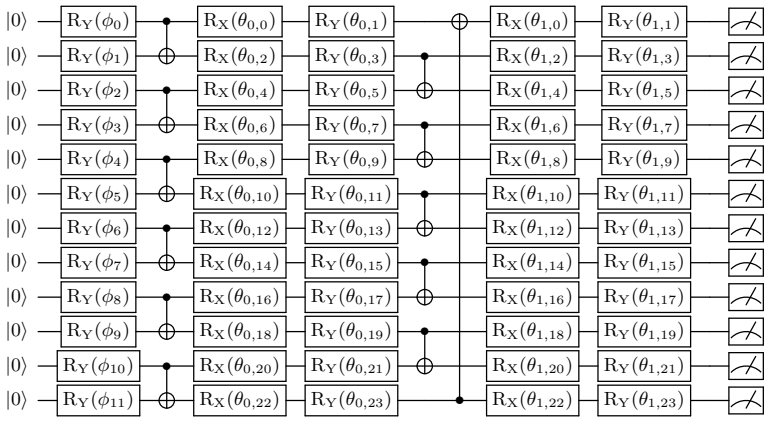


Figure 2. Quantum circuit with parameterized rotation gates. The first layer of gates encodes the data by performing a rotation $\mathrm{R_Y}(\phi_i)$ by the preprocessed data in Figure 1b. ϕ_i are non-trainable. The next two layers entangle the data using two-qubit CNOT operations and learn parameter $\theta_{i,j}$ values that minimize the loss function.

2.2 Classical regression for time-series prediction

To set up a time-series prediction problem using linear regression, suppose we have a time series with n data points, represented as:

$$\{y_t\}_{t=0}^{n-1} = \{y_0, y_1, y_2, \ldots, y_{n-1}\}$$

We want to use a sliding window approach to construct a matrix X and a vector y for supervised learning, where each row in A represents a sequence of past time steps (or "features") used to predict the next time step in y (the "label").

Step 1: Constructing the matrix X and vector y

1. **Choose a window size** m, which determines how many time steps we use to predict the next step.
2. **Form matrix** X: Each row in X is a sequence of m consecutive values from the time series, and there will be $n - m$ rows in total. Formally, $X \in \mathbb{R}^{(n-m)\times m}$ is given by:

$$X = \begin{bmatrix} y_0 & y_1 & \cdots & y_{m-1} \\ y_1 & y_2 & \cdots & y_m \\ y_2 & y_3 & \cdots & y_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n-m-1} & y_{n-m} & \cdots & y_{n-2} \end{bmatrix}$$

3. **Form vector** y: For each row in X, the corresponding entry in y is the next time step in the sequence. Thus, $y \in \mathbb{R}^{(n-m)}$ is given by:

$$y = \begin{bmatrix} y_m \\ y_{m+1} \\ y_{m+2} \\ \vdots \\ y_{n-1} \end{bmatrix}$$

Step 2: Setting Up the Linear System $Xw = y$ The goal is to find a vector $w \in \mathbb{R}^m$ that represents the coefficients used to linearly combine the values in each row of X to predict the corresponding value in y. This setup leads to the linear system:

$$Xw = y,$$

where:

- X is the matrix of input sequences (past values).
- w is the vector of coefficients (weights) that we want to determine.
- y is the vector of target values (next time steps in the time series).

Each element w_i in the vector w represents the weight or influence of the i-th time step in the input window on the prediction of the next time step. Once w is determined, it can be used to make predictions by applying it to new sliding windows from the time series.

Since X is not guaranteed to be invertible, we can set $A = X^T X$ and $b = X^T y$, and use the normal equation $Aw = b$ to solve for w. Vector w that is a solution to this system will minimize the squared error between Xw and y, providing the best linear combination of previous time steps for predicting the next time step.

2.3 Setting for PQC versus classical regression comparison

We compare the performance of PQC and classical regression. All models as loss function use the MSE.

For PQC, we use COBYLA and L-BFGS-B optimizers, each with the maximum number of iterations set to 1000. Quantum circuits use the default Estimator class object for simulation; therefore, circuits run on a noiseless state-vector simulator.

For initial comparison with linear regression, two basic classical models were used - the simple linear model and the simple neural network. Both imitate two layers with 12 parameters on each layer. The only stopping criteria defined is number of epochs equal to 1000. For initial comparison, we did not set seeds.

For further comparison, to ensure more fair conditions and replicability, we extended our experiments. First, we added one more classical model as an improvement for the simple linear model - ridge regression with L2 regularization. The regularization parameter α was tuned via 5-fold cross-validation over $\alpha \in [10^{-6}, 10^{2}]$. Second, we improved the classical neural network - it is trained with L2 weight regularization ($\lambda = 1 \times 10^{-4}$), dropout (0.2), and early stopping (patience = 20) to prevent overfitting. Third, additionally to MSE, we computed mean absolute error (MAE), symmetric

mean absolute percentage error (sMAPE), and mean absolute scaled error (MASE), as well as the number of iterations for each model during the training, and runtime. We also add confidence intervals (bootstrap) for MSE, MAE, and sMAPE. Lastly, we fixed seeds for our experiments to ensure replicability.

2.4 Variational quantum linear regression

In quantum computing, vectors need to be represented as quantum states, which must be normalized. Thus, we need to:

- Normalize the vector b so that it can be represented as a quantum state $|b\rangle$.
- Reformulate the task as finding a $|w\rangle$ (the quantum state corresponding to the vector w) such that $A|w\rangle$ is proportional to $|b\rangle$. This implies that $A|w\rangle = \lambda|b\rangle$ for some scalar λ.

2.4.1 Construction of matrix M and vector b Any matrix M acting on n qubits can be expressed as a linear combination of matrices M_i with coefficients α_i (Reznik and Oppenheim, 2004) as follows:

$$M = \sum_{i=0}^{4^n-1} \alpha_i M_i,$$

where $\alpha_i = \frac{\mathrm{Tr}(M_i \cdot M)}{2^n}$, and M_i is defined as the tensor product of Pauli matrices, $\mathcal{P}_{ij}$, for each qubit:

$$M_i = \bigotimes_{j=0}^{n-1} \mathcal{P}_{ij}.$$

The Pauli matrices are given by:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The key idea is that M can be represented as a linear combination of the matrices M_i, with the basis M_i being all possible combinations of tensor products of the Pauli matrices. Formally, each $\mathcal{P}_{ij}$ is determined by the j-th digit of i in its quaternary (base-4) representation. Specifically, if i is written in base-4 as $i = c_{n-1}c_{n-2}\cdots c_1 c_0$, where $c_j \in \{0, 1, 2, 3\}$, then $\mathcal{P}_{ij} = \sigma_{c_j}$.

However, in reality, we need at most n^2 non-zero terms of M (Werner, 2001).

For instance, for $i = 6$ and $n = 7$, the base-4 representation of i is 0000012. Therefore:

$$\mathcal{P}_{60} = \mathcal{P}_{61} = \mathcal{P}_{62} = \mathcal{P}_{63} = \mathcal{P}_{64} = \sigma_0, \quad \mathcal{P}_{65} = \sigma_1, \quad \mathcal{P}_{66} = \sigma_2.$$

Vector b is normalized to be represented as a quantum state $|b\rangle$. To obtain $|b\rangle$ from b, each element of b should be divided by the square root of the sum of squares of the elements of b. We initialize the quantum state $|b\rangle$ by using Qiskit initialization functionality, which resembles the transpilation of quantum gates.

2.4.2 Reformulating the Task as Finding a State Vector $|w\rangle$ As discussed in Section 2.2, for input X and y, we aim to solve the equation for the vector w:

$$Aw = b,$$

where $A = X^T X$ and $b = X^T y$. In the quantum setting, the solution for w is typically found in terms of a unit vector:

$$M\,|w\rangle = |b\rangle\,,$$

where $|w\rangle = \frac{w}{\|w\|}$ and $|b\rangle = \frac{b}{\|b\|}$, and M represents the matrix decomposition of A in terms of Pauli operators, i.e., $M = \sum_i \alpha_i M_i$.

The problem is that we cannot directly compute w from the quantum state $|w\rangle$, as the algorithm only provides the unit vector $|w\rangle$, and the norm $\|w\|$ is unknown. Although the vector $|w\rangle$ will be solved by VQLS using the decomposition of A, to find w, we can rescale the answer $|w\rangle$ using the following relation

$$\frac{A\,|w\rangle}{\|A\,|w\rangle\,\|} \approx |b\rangle$$

Multiplying both sides of this equation by $\|b\|$, we get b on the right side:

$$A \cdot \underbrace{\frac{\|b\|}{\|A\,|w\rangle\,\|}\,|w\rangle}_{w} \approx b$$

Thus, the solution for w is:

$$w \approx \frac{\|b\|}{\|A\,|w\rangle\,\|}\,|w\rangle$$

,where $|w\rangle$ is the result obtained from the VQLS and A input matrix.

2.4.3 Parameterized ansatz We did VQLS experiments with 2, 3, and 4 qubits, that represent time series of length n of 4, 9, and 16. For each size of experiment, we used the corresponding ansatz. An ansatz is a PQC, whose parameters can be adjusted to generate a wide range of different quantum transformations and quantum states. An example of an ansatz for 2 qubits is depicted on Figure 3.

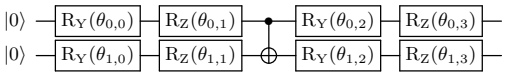


Figure 3. Quantum circuit for parameterized ansatz with 2 qubits ($n = 4$).

2.4.4 Cost Function The goal of the VQLS is to find the optimal parameter vector φ for the ansatz state $|\psi\rangle = M\,|x(\varphi)\rangle$, where $|x(\varphi)\rangle$ is the state after the ansatz and M decomposition of input matrix. We define the cost function similarly to the one in (Bravo-Prieto et al., 2023), but with a slight modification

$$C = \langle\psi|\psi\rangle - \langle\psi|b\rangle\langle b|\psi\rangle$$

Since $|\psi\rangle$ is a unit vector, the first term simplifies to 1, yielding:

$$C = 1 - |\langle b|\psi\rangle|^2$$

This expression can be rewritten as:

$$\langle b|\psi\rangle = \langle b|M\,|x(\varphi)\rangle\rangle = \langle b|\sum_i \alpha_i M_i|x(\varphi)\rangle = \sum_i \alpha_i \langle b|M_i|x(\varphi)\rangle.$$

To evaluate each term in the sum, we use the Hadamard test. This means that the cost function is computed as a sum over all M_i, where for each term, the following procedure is performed: first, the ansatz state $|x(\varphi)\rangle$ is prepared; then, the corresponding M_i transformation is applied. Afterward, the transformation that prepares the unit vector b is applied to the state, and the resulting state is used as input for the Hadamard test as in Figure 4. The Hadamard test allows us to evaluate each term of $\langle b|\psi\rangle$. Finally, the result of the Hadamard test is multiplied by the corresponding α_i, and the final cost function is obtained by summing over all terms.

As an alternative to the Hadamard test, the Swap test can be used. The approach is similar to the Hadamard test, and our experiments with the Swap test have shown the same outcomes as with the Hadamard test, so we omit the details of the Swap test implementation.

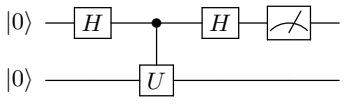


Figure 4. Hadamard Test circuit

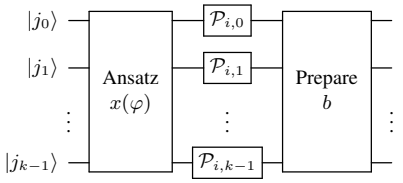


Figure 5. Quantum Circuit U for the Hadamard Test, shown in Figure 4.

The circuit in Figure 5 illustrates the estimation of $\langle b|M_i|x(\varphi)\rangle$ for a matrix M_i, which represents the i-th term in the decomposition of M of Pauli gates $\mathcal{P}_0, \ldots, \mathcal{P}_{k-1}$. Initially, all $k = \log n$ qubits are prepared in the state $|0\rangle$. A parameterized ansatz $x(\varphi)$ is then applied, followed by the application of the M_i decomposition into Pauli gates $\mathcal{P}$. Finally, a transformation is applied to prepare the vector b.

3 Experimental results

In this section, we present the results of our experiments, including training and test performance of quantum and classical models, learning dynamics, and error comparisons.

3.1 First comparison of performance of PQC and classical regression

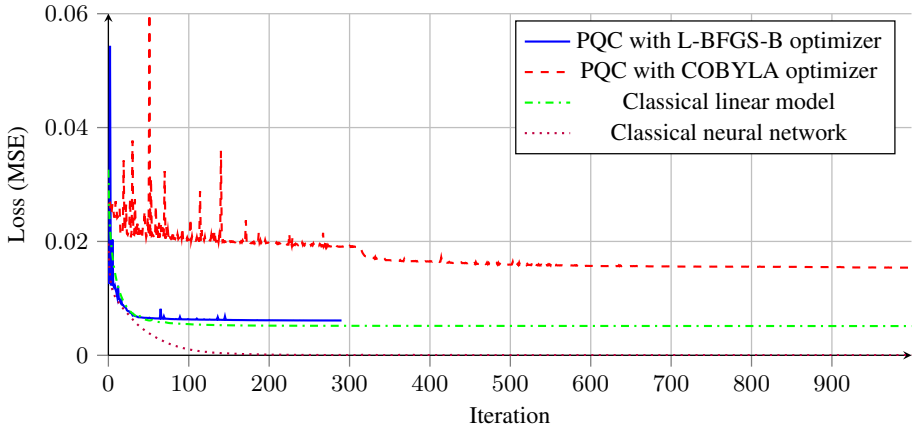


Figure 6. Loss function (MSE) comparison of different optimization methods and models over iterations. PQC with L-BFGS-B optimizer converged after 291 iterations.

Figure 6 demonstrates the learning curve of different models, showing how the training dataset MSE evolved over iterations. The PQC model using the L-BFGS-B optimizer showed the fastest convergence, reaching its minimum error at 291 iterations. In contrast, the COBYLA optimizer required more iterations and converged to a slightly higher error level. Among the classical models, the linear model achieved rapid convergence but displayed a higher final error compared to the quantum model with L-BFGS-B. The neural network showed fast improvement over iterations, leading to overfitting, which is clearly visible in Figure 7. However, the PQC with L-BFGS-B outperformed all the classical models in terms of test set MSE, as summarized in Table 1, indicating a potential advantage in using quantum-enhanced optimizers for this problem.

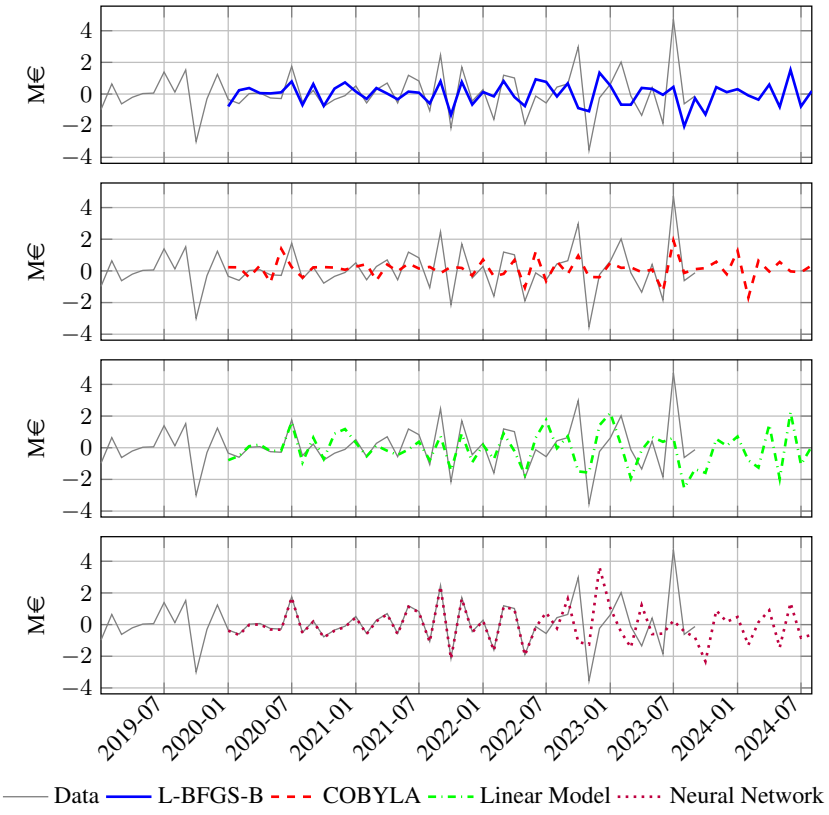


Figure 7. Comparison of different model predictions of the month-on-month sales growth against the actual data.

Table 1. MSE of models on training and test sets

Model	**Training Set MSE**	**Test Set MSE**
PQC with COBYLA	0.01257	0.02106
PQC with L-BFGS-B	0.00612	0.04418
Classical Linear Model	0.00514	0.05177
Classical Neural Network	0.00003	0.05767

The final MSE values on the test set, summarized in Table 1, show that the PQC with COBYLA achieved the best performance on the test set, followed by the PQC with L-BFGS-B. However, both classical models were able to fit the training data better, and with proper regularization would likely achieve improved test MSE. No explicit regularization was applied to any of the models. The neural network, in particular, exhibited signs of overfitting, suggesting that techniques such as early stopping or other forms of regularization could significantly improve its performance.

3.2 Further comparison of performance of PQC and classical regression

In this subsection, we show the results after performing experiments for a fairer and detailed model comparison, described in the last paragraph of Section 2.3. Results in Table 2 and Table 3 were achieved with the random seed set to 25. We repeated our experiments with different fixed seeds (96, 123, 817, 4321, 9999), and the results were similar to the ones that we obtained with the seed set to 25.

Table 2. Evaluation metrics for all models on training and test sets. The lowest Test MSE is shown in bold. MASE is reported once per model.

Model	Dataset	MSE	MAE	sMAPE	MASE	Epochs/Fits	Runtime (s)
PQC (COBYLA)	Test	**0.02065**	0.10191	0.475			
PQC (COBYLA)	Train	0.01276	0.08883	0.533	0.692	1000	807.7
PQC (L-BFGS-B)	Test	0.04363	0.15779	0.636			
PQC (L-BFGS-B)	Train	0.00614	0.06432	0.481	1.072	264	9511.5
Linear Model	Test	0.05111	0.19221	0.676			
Linear Model	Train	0.00516	0.05417	0.449	1.306	1000	4.4
Ridge Regression	Test	0.04211	0.15167	0.651			
Ridge Regression	Train	0.00717	0.06725	0.487	1.031	100	0.2
Neural Network	Test	0.05600	0.19686	0.707			
Neural Network	Train	0.00712	0.06336	0.445	1.338	66	22.4

Table 2 summarizes all evaluation metrics for training and testing datasets. The column *Epochs/Fits* for PQC shows the number of optimizer iterations before the algorithm has stopped, and for classical models, the column shows the number of epochs

before the algorithm has stopped. The *Runtime* column shows the total training time of a given model.

As we can see, PQC with COBYLA still achieved the best performance on the test dataset, among all error measures. We can also see that PQC with L-BFGS-B took the longest time, taking more than 30 seconds per optimization iteration. The performance of PQC with L-BFGS-B and ridge regression is similar across all evaluation metrics.

We can also see in Table 2 that results for neural network have slightly improved, but for some fixed random seeds, neural network showed better results than other classical counterparts, and sometimes even surpasses PQC with L-BFGS-B (e.g., for seed 817). In most cases, higher error rates on the training set resulted in lower error rates on the test set, but not without exceptions (e.g., for seed 4321).

Table 3. Evaluation metrics with 95% confidence intervals for all models on training and test sets.

Model	Dataset	MSE [95% CI]	MAE [95% CI]	sMAPE [95% CI]
PQC (COBYLA)	Test	0.020 (0.006–0.039)	0.101 (0.053–0.157)	0.475 (0.304–0.651)
PQC (COBYLA)	Train	0.013 (0.007–0.020)	0.088 (0.065–0.114)	0.530 (0.414–0.651)
PQC (L-BFGS-B)	Test	0.043 (0.017–0.076)	0.157 (0.092–0.226)	0.637 (0.440–0.833)
PQC (L-BFGS-B)	Train	0.006 (0.004–0.009)	0.064 (0.048–0.080)	0.482 (0.372–0.594)
Linear Model	Test	0.051 (0.025–0.083)	0.191 (0.133–0.252)	0.676 (0.496–0.845)
Linear Model	Train	0.005 (0.003–0.008)	0.054 (0.038–0.071)	0.449 (0.334–0.572)
Ridge Regression	Test	0.041 (0.014–0.074)	0.150 (0.085–0.225)	0.652 (0.454–0.836)
Ridge Regression	Train	0.007 (0.004–0.011)	0.067 (0.049–0.086)	0.486 (0.381–0.601)
Neural Network	Test	0.055 (0.026–0.091)	0.196 (0.133–0.264)	0.707 (0.537–0.850)
Neural Network	Train	0.007 (0.003–0.012)	0.063 (0.044–0.085)	0.445 (0.332–0.563)

Table 3 shows 95% confidence intervals (bootstrap), which are related to the evaluation metrics from Table 2.

Experiments of this subsection were performed on a laptop, equipped with 16 GB of RAM and an Intel Code i5-7300U CPU (no GPU), running Windows 10. We had the following versions of libraries during experiments: 'python_version': '3.8.8', 'qiskit_version': '1.1.1', 'qiskit_machine_learning': 'qiskit_machine_learning 0.7.2'. Results with other seeds are available in our code repository: `https://github.com/infenrio/pqc_and_vqls`. The code also contains seed setting, data preprocessing, optimizer settings, and circuits.

3.3 Performance of VQLS

Testing the VQLS algorithm with the provided matrix decomposition from (Reznik and Oppenheim, 2004) yielded unsatisfactory results. The time-series forecasting predictions were indistinguishable from random values, offering no reasonable forecast. This was also observed when we tested the approach on synthetic data from $Forse\ AI$. One key issue is the extremely high condition number of the corresponding arbitrary matrix.

Transforming a custom matrix with n rows and columns requires $O(n^2)$ Pauli matrices (Werner, 2001).

An alternative solution for addressing this issue in future work could involve applying the Gram-Schmidt orthogonalization procedure. Another potential factor contributing to the poor performance is the COBYLA optimizer, which demands a large number of iterations to converge.

On the other hand, results were satisfactory when we tried time series that strictly grow or strictly decline. In both cases, the output continued the growing/declining trend. This result may have potential uses in trading and investment areas in cases of growing and declining trends. Interestingly, we achieved ideal results when the data represented a geometric progression (both increasing and decreasing), but it is quite an artificial case to be considered useful practically.

Overall, we think that this approach deserves further investigation, as it is based on shallow quantum circuits on a small number of qubits, which is within the capacity of available quantum computers.

4 Conclusion

In this section, we conclude the paper by summarizing the key findings, reflecting on their potential impact, and mentioning limitations and future research directions.

Our findings can be situated within the broader landscape of quantum approaches to time-series forecasting. Several studies have explored shallow PQCs with data reuploading strategies, which enhance expressivity without deep architecture and have shown promising results for sequential learning tasks. Recent work has adapted it to forecasting tasks (e.g., for traffic data) with promising results (Pérez-Salinas et al., 2020). Another promising direction is quantum reservoir computing, where fixed random quantum dynamics are used as a nonlinear feature map. Such models have been shown effective for chaotic time-series, stock market prediction, and system memory tasks (Kutvonen et al., 2020). Alternatively, hybrid architectures such as quantum long short-term memory networks (QLSTM) (Khan et al., 2024), provide a natural way to combine recurrent inductive biases with quantum feature maps. Compared to these approaches, our PQC-based regressor emphasizes shallow, directly trainable circuits optimized for forecasting tasks. PQC as architecture represents a middle ground: more structured than fixed-reservoir QRC, but simpler than full hybrid RNNs.

The considered PQC and VQLS models may scale well. For the PQC model, the window lag is equal to the number of qubits; therefore, if PQCs with shallow depth can show consistent results for a large number of qubits, it can be practical on quantum computers with a high number of qubits and a moderate gate error rate. For the VQLS model, the window lag is proportional to the square of the number of qubits. VQLS can become practical on intermediate-scale quantum computers with low gate error rates.

Our findings demonstrate that hybrid quantum-classical approaches can achieve forecasting accuracy comparable to simple classical models, highlighting their potential role in practical forecasting solutions when quantum hardware improves. Overall, our results suggest that the hybrid quantum-classical approach shows promise, but careful

tuning of hyperparameters, such as the choice of optimizer, plays a crucial role and requires extensive experimentation. Once effective quantum models are identified, they could be integrated into ensembles of models, potentially aiding with regularization and complementing classical models by offering unique strengths and weaknesses, thereby enriching the ensemble's overall performance.

We must acknowledge several limitations of our study. First, we conducted experiments on synthetic data, which simplifies many challenges of real-world forecasting; thus, the generalizability of our claims to diverse application domains remains untested. Second, we evaluated only relatively shallow circuits and small datasets, and the experiments were limited to noiseless simulators. Third, the comparison to classical baselines can be strengthened further, e.g., through systematic hyperparameter tuning and inclusion of more expressive models.

Future work should focus on extending experiments to real-world datasets (including diverse real benchmarks and multiple series), exploring noise-resilient and data reuploading ansätze, incorporating regularization for fairer comparisons, and evaluating scalability with respect to circuit depth and qubit count. An interesting future direction would be to integrate PQC-based forecasters with classical models – this may further clarify their unique contributions to time-series prediction.

Acknowledgements

The authors would like to express their gratitude to *Forse AI* for providing the synthetic dataset used in this study and for their financial support, which made this research possible.

The authors thank anonymous reviewers for their very helpful comments.

References

Benedetti, M. et al. (2019). Parameterized quantum circuits as machine learning models, *Quantum Science and Technology* **4**(4) (2019), p. 043001.

Bravo-Prieto, C. et al. (2023). Variational Quantum Linear Solver, *Quantum* **7** (Nov. 2023), p. 1188. DOI: 10.22331/q-2023-11-22-1188.

Fisher, A. R., Hodgdon, T. S., Lewis, M. G. (2024). Time-Series Forecasting Methods: A Review, ERDC/GRL TR-24-3. DOI: 10.21079/11681/49450.

Halevy, A. et. al. (2009) The Unreasonable Effectiveness of Data *IEEE Intelligent Systems* **24**(2) (2009), p. 8. DOI: 10.1109/MIS.2009.36.

Harrow, A. W., Hassidim, A, Lloyd, S. (2009). Quantum Algorithm for Linear Systems of Equations, *Phys. Rev. Lett.* **103** (15 Oct. 2009), p. 150502. DOI: 10.1103/PhysRevLett.103.150502.

Khan, S. Z. et al. (2024). Quantum long short-term memory (QLSTM) vs. classical LSTM in time series forecasting: a comparative study in solar power forecasting, *Front. Phys.* **12** (Oct. 2024). DOI: 10.3389/fphy.2024.1439180.

Kutvonen, A., Fujii, K., Sagawa, T (2020). Optimizing a quantum reservoir computer for time series prediction, *Sci Rep.* **10**(1) (Sep. 2020), p. 14687. DOI: 10.1038/s41598-020-71673-9.

Lara-Benítez, P., Carranza-García, M., Riquelme, J. C. (2021). An Experimental Review on Deep Learning Architectures for Time Series Forecasting, *International Journal of Neural Systems* **31**(03) (2021). DOI: 10.1142/S0129065721300011.

Masini, R. P., Medeiros, M. C., Mendes, E. F. (2020). Machine Learning Advances for Time Series Forecasting, *Journal of Economic Surveys* **37**(1) (Feb. 2023), pp. 76-111. DOI: 10.1111/joes.12429.

Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J. I. (2020). Data re-uploading for a universal quantum classifier, *Quantum* **4** (Feb. 2020), p. 226. DOI: 10.22331/q-2020-02-06-226.

Preskill, J. (2018). Quantum Computing in the NISQ era and beyond, *Quantum* **2** (Aug. 2018), p. 79. DOI: 10.22331/q-2018-08-06-79.

Reznik, B., Oppenheim, J. (2004). A probabilistic and information theoretic interpretation of quantum evolutions, *Quantum Physics* (Aug. 2004). DOI: 10.1103/PhysRevA.70.022312.

Werner, R. F. (2001). All teleportation and dense coding schemes, *Journal of Physics A: Mathematical and General* (2001). DOI: 10.1088/0305-4470/34/35/332.

Zhang, C. et. al. (2021) Understanding deep learning (still) requires rethinking generalization, *Communications of the ACM* **64**(3) (2021), p. 107. DOI: 10.1145/3446776.

Received June 19, 2025 , revised January 18, 2026, accepted February 18, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 190-201
https://doi.org/10.22364/bjmc.2026.14.1.10

Uncertainty of a State Vertical Reference System below 1 mgpu – Is It Possible?

Vasil CVETKOV, Christina MICKRENSKA

University of Architecture, Civil Engineering and Geodesy,
Hristo Smirnenski Blvd. 1, Sofia, Bulgaria

tzvetkov_vasil@abv.bg, kmikrenska_fgs@uacg.bg

ORCID 0000-0001-9628-6768, ORCID 0000-0002-7054-1799

Abstract. Precise geometric levelling has been the most accurate method for measuring height differences on Earth for over 150 years. Yet, its accuracy has seen little improvement in the past 80 years. This stagnation is mainly due to the rise of GNSS technologies and a lack of progress in levelling data processing. This article presents new statistical insights into how levelling uncertainties accumulate and introduces a revised algorithm for adjusting levelling networks. Testing the method with data from Finland's Third Levelling (1978–2006) under 66% showed standard deviations of benchmark geopotential numbers below 0.7 mgpu (or 0.7 mm). These results prove that, by rethinking outdated approaches, a fifteenfold increase in accuracy is achievable.

Keywords: Probability Theory, Reference Systems, Geometric Levelling, Vertical Measurements, Adjustment Algorithms

1. Introduction

1.1. Precise levelling classical assumptions

According to Kääriäinen (Kääriäinen, 1966, pp. 52), "the errors are divided into two classes, accidental and systematic, which are assumed to be independent of each other. The accidental errors are caused by sources that are independent in all successive observations and obey the laws of Gauss. The systematic errors are due to causes acting similarly on successive or adjacent levelling observations; these do not obey the laws of Gauss. They become accidental only for distances exceeding a certain limit Z of the order of several tens of kilometres. The total error is the combined influence of both types of error and is calculated for distances longer than the above limit as the quadratic sum of the accidental and systematic errors". This popular belief about the accumulation of levelling uncertainties dominates in many scientific studies (Gerlach and Rummel, 2024; Izvoltova et al., 2024; Kääriäinen, 1966; Lyszkowicz and Leonczyk, 2006; Saaranen et al., 2021; among others) where levelling data is processed. However, it is not supported by some, and likely not by any, of the levelling data.

1.2. Statistical relationship between line discrepancies and their length

According to (Cvetkov, 2024g), the coefficients of correlation R of line descripencies $|D|$ to the square root of line length $\sqrt{L}$ are 0.530, 0.456, 0.530, and 0.444, respectively for the Second Levelling of Bulgaria /1953-1957/, the Third Levelling of Bulgaria /1975-1984/, The Second Levelling of Finland /1935-1955/, the Third Levelling of Finland /1978-2006/. According to (Lyszkowicz and Leonczyk, 2006, Figure 7), the coefficient of correlation between line length L and line discrepancies $|D|$ in the case of the Polish Fourth Levelling data (1999-2002) is R = 0.41. Taking into account that the correlation between $\sqrt{L}$ and L is usually higher than 0.99, it follows that the coefficient of correlation between the line discrepancies $|D|$ and the square root of line length , for the Polish Fourth network will be in the same magnitude, or R ≈ 0.41. Thus, for the coefficients of determination R^2 of descripencies $|D|$ to the square root of line length $\sqrt{L}$ for the Second Levelling of Bulgaria /1953-1957/, the Third Levelling of Bulgaria /1975-1984/, The Second Levelling of Finland /1935-1955/, the Third Levelling of Finland /1978-2006/, and the Fourth Levelling of Poland are 0.281, 0.208, 0.281, 0.197 and 0.17. This simply shows that the variations in the line discrepancies D can be explained up to 30% by the square root of the levelling line length $\sqrt{L}$. This conclusion is also valid for the explanation of D by L. A graphical evidence of this fact is the dispersion of line discrepancies |D| concerning $\sqrt{L}$ for the data of the Third Levelling of Finland.

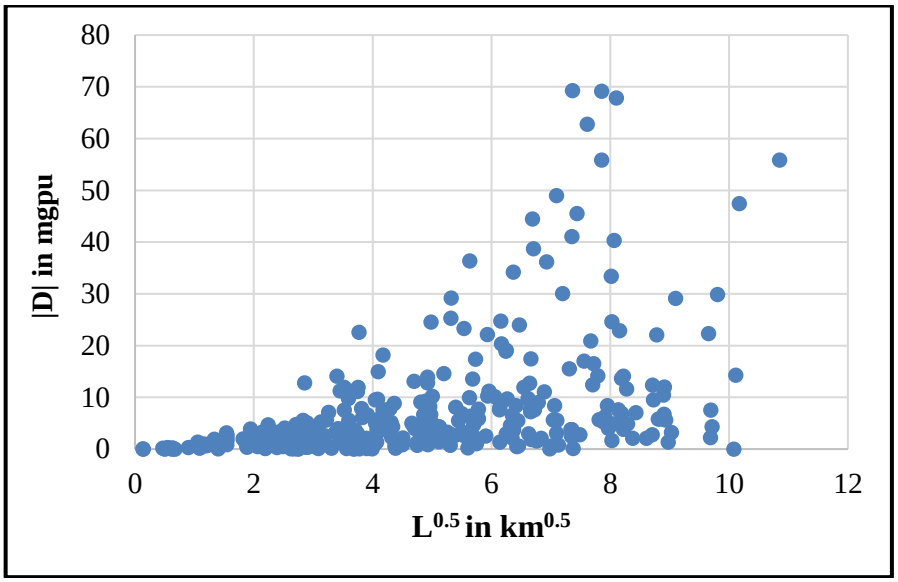


Fig. 1. Dispersion of line discrepancies |D| concerning $\sqrt{L}$ for the data of the Third Levelling of Finland. The data are taken from (Saaranen et al., 2021)

A similar pattern of spread of |D| in respect of $\sqrt{L}$ (L) is presented in (Cvetkov, 2023; Cvetkov, 2022; Lyszkowicz and Leonczyk, 2006) for the Second Levelling of Finland, the Third Levelling of Bulgaria, and the Fourth Levelling of Poland. Taking into account

the differences in measurement epochs, levelling equipment, methodology of performance, terrain and climate conditions in Finland, Poland and Bulgaria, it will be naive to say that this common pattern of the spread of |D| in respect of $\sqrt{L}$ (L) is incidental. On the contrary, the pattern is natural for precise geometric levelling, and its physical explanation will be given in a future publication.

1.3. Statistical relationship between loop closing errors and loop circumferences

Let us explore the relationship between loop closing errors $|W|$ and loop circumferences L or its square root $\sqrt{L}$. According to (Cvetkov, 2024a; Cvetkov, 2024g), the coefficients of correlation R of closing errors $|W|$ to the square root of line length $\sqrt{L}$ are 0.447, 0.348, 0.234, and -0.006, respectively for the Second Levelling of Bulgaria /1953-1957/, the Third Levelling of Bulgaria /1975-1984/, The Second Levelling of Finland /1935-1955/, the Third Levelling of Finland /1978-2006/. According to (Lyszkowicz and Leonczyk, 2006, Figure 8), the coefficient of correlation between loop circumference L and loop closing errors $|W|$ in the case of the Polish Fourth Levelling data (1999-2002) is R = 0.12. Thus, the coefficient of correlation between the loop closing errors $|W|$ to the square root of the loop circumference, for the Polish Fourth network, will be R ≈ 0.12. Thus, for the coefficients of determination R^2 of $|W|$ to the square root of loop circumference $\sqrt{L}$ for the Second Levelling of Bulgaria /1953-1957/, the Third Levelling of Bulgaria /1975-1984/, The Second Levelling of Finland /1935-1955/, the Third Levelling of Finland /1978-2006/, and the Fourth Levelling of Poland are 0.200, 0.121, 0.055, 0.000 and 0.014. This simply shows that the variations in the closing errors W can be explained up to 20% by the square root of the loop length $\sqrt{L}$. This conclusion is also valid for the explanation of W by L. A graphical evidence of this fact is the dispersion of the $|W|$ concerning $\sqrt{L}$ for the data of the Second and the Third Levelling of Finland given by Figures 2 and 3.

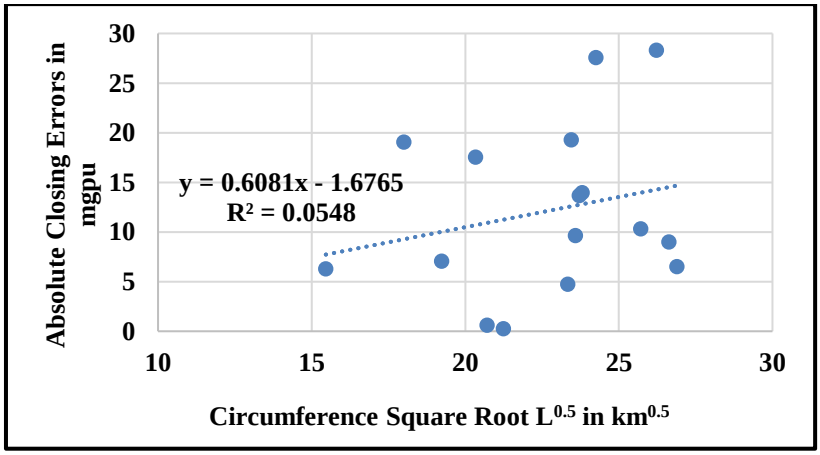


Fig. 2. Closing Errors in the Second Levelling of Finland.
The data are taken from (Kääriäinen, 1966).

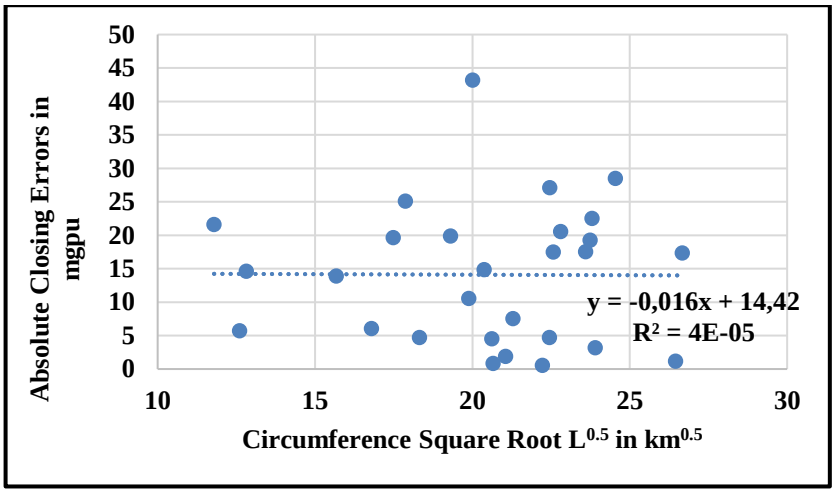


Fig. 3. Closing Errors in the Third Levelling of Finland.
The data are taken from (Saaranen et al., 2021)

According to Figures 2 and 3, which present the determination R^2 and the trend of the increase of absolute closing errors |W| with respect to the squared loop length $\sqrt{L}$ in km, there is a weak (almost nothing) relationship between |W| and $\sqrt{L}$. A similar pattern of the spread of $|W|$ with respect to $\sqrt{L}$ (L) is presented in (Cvetkov, 2022) for the Third Levelling of Bulgaria. Again, given the differences in measurement epochs, levelling equipment, procedures, and the terrain and climate conditions across Finland, Poland, and Bulgaria, it would be naive to claim that the common pattern in the spread of $|W|$ concerning $\sqrt{L}$ (or L) is merely coincidental. On the contrary, this pattern is inherent to precise geometric levelling, and its physical explanation will be presented in a future publication.

1.4. Observations true error vs. observation average true error

The probability *Cv(n)* that the average of *n* independent random variables is closer to the distribution expectation than some of the independent variables is a function of the entropy of the standard Normal distribution and the number of variables *n*. The probability *Cv(n)* can be given by equation (1), where the coefficients *a*, *b*, and *c* depend on the distribution and the number of variables *n* (Cvetkov, 2024b, 2024c).

$$Cv(n) = 0.25.log_b(2\pi e).(n^{-c} - n^{-a}) \quad (1)$$

Based on (1), it can be calculated that in the case of two measurements, i.e., $n = 2$, the probability *Cv(2)* tends to 30% if both measurements are normally distributed. In our little experiment, we obtained *Cv(2)* = 32%. If one generates their random sample, the result will differ, but if their sample size is $n \to \infty$, then their $Cv(2) \to 30\%$.

Suppose the situation where both measurements derive from the distribution with the highest entropy, i.e., the Uniform distribution, then $Cv(2) \rightarrow 33\%$. Thus, in more than 66% of the cases, the value of one of the two measurements is closer to the true value of the measured quantity in comparison to the mean. Figure 5 shows how often the first observation, the second observation, or their average comes closest to a known expected value across different distributions (Cvetkov, 2024e).

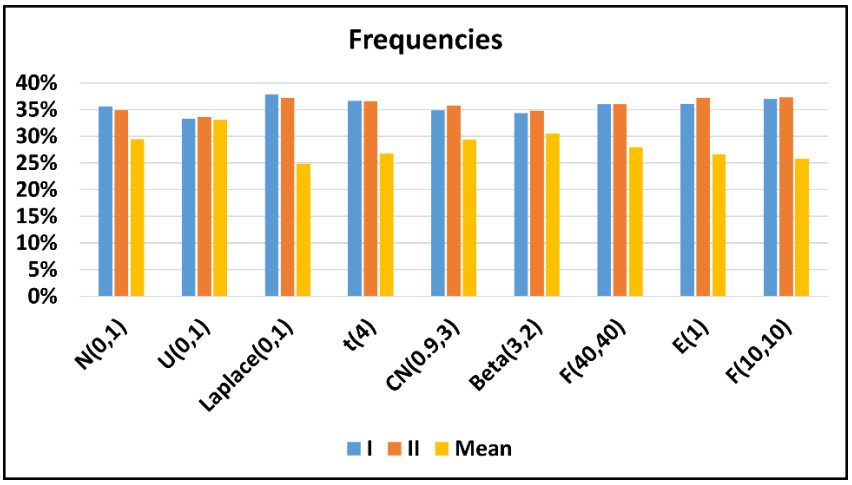


Fig. 5. The frequencies of occurrence of the first, the second observation, or their mean are closest to a known expectation of the applied distributions.

Figure 5 shows that for skewed (asymmetrical) distributions, the average of two observations is more accurate than each observation in less than 30% of the cases. We should remember that with only two real observations, we can not determine their distribution, so assuming normality is just a convenient simplification.

1.5. Research objectives

Building on the statistical findings and probability insights presented above, the primary objective of this article is to expand on Cvetkov's study (Cvetkov, 2024d) by applying newly proposed weighting methods and reconfiguring the levelling network as a free network, with the datum point set at its centroid. A secondary goal is to assess how each step of the proposed algorithm contributes to improved accuracy and to identify the causes behind systematic tilts observed in certain reference systems (e.g., Gerlach and Rummel, 2024).

2. Data and data processing

2.1. Data processing

In this study, we used data from the Third Levelling of Finland (Saaranen et al., 2021), previously analysed by Cvetkov (Cvetkov, 2024d). The network layout is shown in Figure 6, and the data used are listed in Table 1.

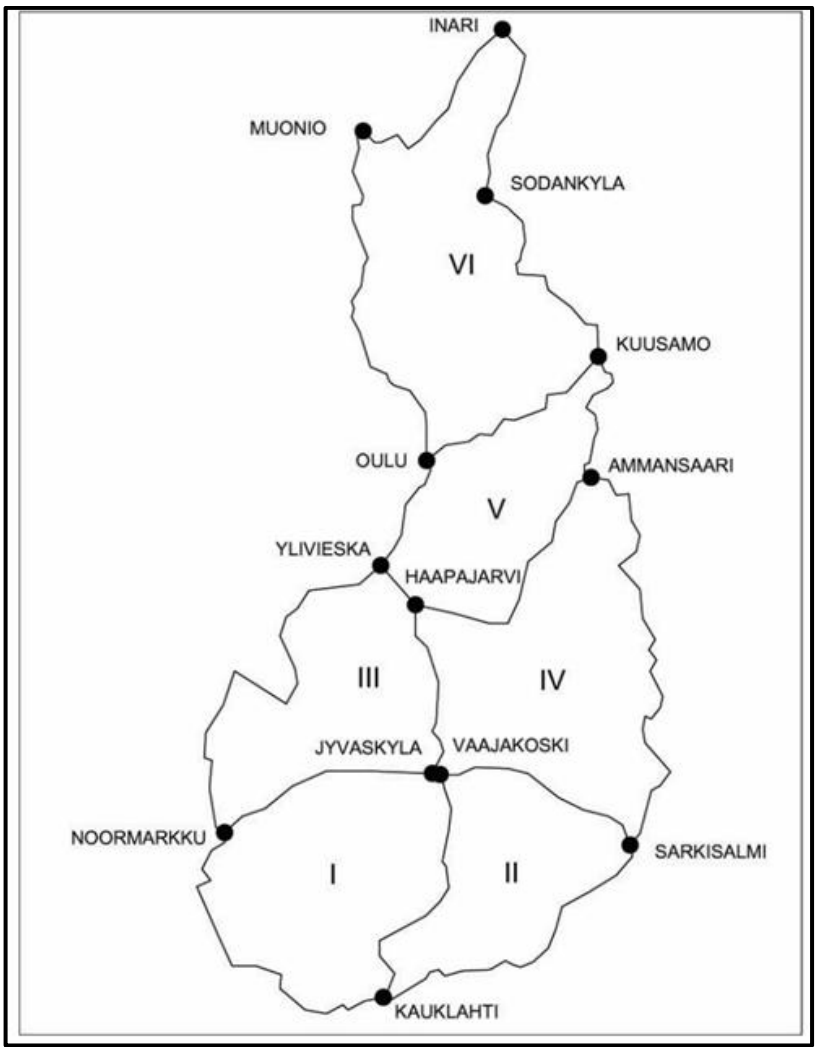


Fig. 6. Scheme of the analysed network, part of the Third Level of Finland network (Cvetkov, 2024d; Saaranen et al., 2021).

Table 1. Summarised data about levelling lines, their length, differences *I* and *II* between geopotential numbers of the start and the end line benchmarks, and the averages of both measurements *I* and *II*.

Line	Distance	Height differences*		Mean
		I	***II***	$(\boldsymbol{I}+\boldsymbol{II})/\mathbf{2}$
	(km)	**(mgpu)**	**(mgpu)**	**(mgpu)**
Kauklahti-Noormarkku	363.778	35284.13	35231.27	**35257.70**
Noormarkku-Jyväskylä	257.280	42172.85	**42164.91**	42168.88
Kauklahti-Vaajakoski	337.837	**80998.47**	81017.91	81008.19
Jyväskylä-Vaajakoski	4.836	**3580.65**	3581.17	3580.91
Vaajakoski-Särkisalmi	239.680	-16433.88	**-16399.66**	-16416.77
Kauklahti-Särkisalmi	397.821	64603.73	64598.05	**64600.89**
Noormarkku-Ylivieska	490.659	21623.15	**21533.81**	21578.48
Haapajärvi-Ylivieska	64.685	-40398.99	-40415.51	**-40407.25**
Jyväskylä-Haapajärvi	201.790	**19778.81**	19749.75	19764.28
Haapajärvi-Ammänsaari	302.491	94304.50	**94322.86**	94313.68
Sarkisalmi-Ammänsaari	513.336	126924.65	**126920.11**	126922.38
Ylivieska-Oulu	129.545	**-57206.60**	-57204.58	-57205.59
Oulu-Kuusamo	254.097	**257863.35**	257827.31	257845.33
Ammänsaari-Kuusamo	173.309	**65926.36**	65930.36	65928.36
Oulu-Muonio	434.562	237884.45	237875.21	**237879.83**
Kuusamo-Sodankylä	266.298	-91600.12	-91656.04	**-91628.08**
Sodankylä-Inari	210.199	-54031.64	-54184.14	**-54107.89**
Muonio-Inari	255.329	**-125752.04**	-125707.40	-125729.72

*The values of the *I* and the *II* measurements we calculated based on the information given by Appendix C, columns 5-11 in the study by Saaranen et al. (Saaranen et al., 2021). All height differences contain rod metre, refraction, and temporal tidal and land uplift corrections.

2.2. Data processing

The basic steps of the proposed adjustment of the analysed part of the Third Levelling of Finland network (1978–2006) are as follows:

- Step 1 – In our network, we have 18 lines and three elevation values for each line, i.e., from I and II measurements and their average. Thus, in our case, we have 3^{18} = 387,420,489 combinations between line elevations, or 3^{18} independent

adjustments. Using those 3^{18} independent adjustments, we found those line elevations that minimised the loop closing errors in the network shown in Figure 6. These selected values are highlighted in bold in Table 1.

- Step 2 - Next, the weights of the selected line elevations were obtained based on the real impact of each one on the final network accuracy. The network was readjusted 18 times, each time omitting a different levelling line. For each case, the network was adjusted using the remaining 17 lines, and the standard errors of the adjusted benchmark heights were summed. If a levelling line degrades the network accuracy, omitting it leads to a smaller sum of nodal benchmark standard errors, implying a lower weight for that line. Because the number of levelling lines is the same in each independent adjustment, the average of the nodal benchmark standard errors (ASE) was used instead of their sum. From the 18 obtained ASE values, their mean value (ASE_{MEAN}) was calculated, and the weight w_{L_i} for line *i* was then computed using equation (2):

$$w_{L_i} = (ASE_i/ASE_{MEAN})^2 \quad (2)$$

- Step 3 - Finally, the network was adjusted as a free network using the selected height differences for each line together with the assumption-free weights defined by equation (2).

3. Results

Figure 7 compares the accuracy of the classical adjustment method (Saaranen et al., 2021) with the individual steps outlined in Section 2.2. Figure 8 presents the differences between the official benchmark geopotential numbers and those derived from our adjustment approach.

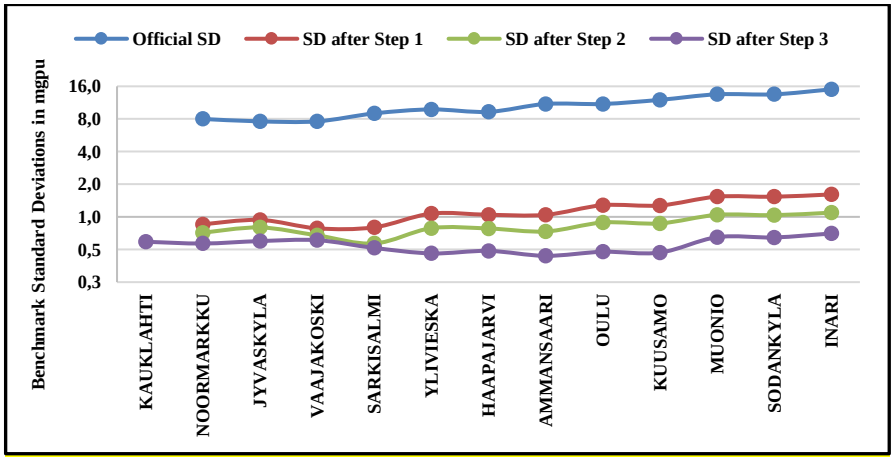


Fig. 7. Standard deviations of the adjusted benchmark geopotential numbers in mgpu – official results (blue), after data processing Step 1 (red), Step 2 (green), and Step 3 (violet).

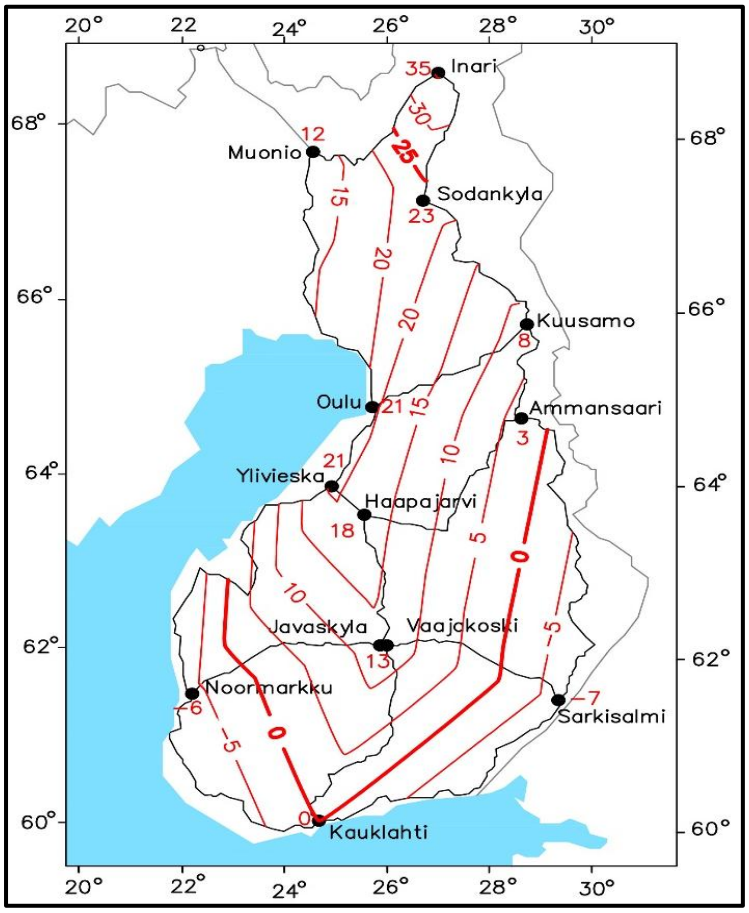


Fig. 8. Differences in mgpu between the official benchmark geopotential numbers (Saaranen et al., 2021) and those calculated using our adjustment method.

4. Discussion

According to Figure 7, the standard deviations of all adjusted benchmark geopotential numbers, obtained after Step 3, are below 0.75 mgpu. The adjusted geopotential number of the Ammansaari benchmark has a minimal standard deviation of 0.44 mgpu. This benchmark is close to the network mass centre. The benchmarks in the network's periphery, e.g., Inari, Sodankyla, and Mounio, have greater standard deviations of the adjusted geopotential numbers, respectively 0.70, 0.65, and 0.65 mgpu. The mean of the standard deviations is 0.55 mgpu. Comparison between the standard deviations of the adjusted geopotential numbers pictured in Figure 7 and the standard deviations of the adjusted geopotential number of the same benchmarks taken from the Saaranen et al.'s

study (Saaranen et al., 2021, Figure 6.3) shows that the uncertainty of a levelling network, adjusted in the manner presented here, can be reduced more than 15-20 times.

Additionally, Figure 7 illustrates a stepwise improvement in the accuracy and homogeneity of the adjusted network. A two-tailed paired T-test comparing the mean standard deviations of the benchmark-adjusted geopotential numbers yielded p-values of 1.28×10^{-8}, 1.23×10^{-5}, and 0.015 for the comparisons between the official results and Step 1, Step 2, and Step 3 given in Section 2.2, respectively. These results confirm that each step produced a statistically significant accuracy gain at a confidence level higher than 95%. Notably, the greatest improvement in adjustment accuracy comes from Step 1.

A key aspect of levelling network adjustment is the reliability of the final results. While Least Squares Estimates (LSE) are known to be the Best Linear Unbiased Estimates (BLUE) or Minimum Variance Unbiased Estimates (MVUE) under the assumption of normally distributed measurement errors, they lack robustness against outliers and systematic errors. Instead of isolating such anomalies, LSE spreads its influence across all observations, leading to potential degradation of results (Duchnowski and Wyszkowska, 2019; Javadi, 2017; among others).

One source of this degradation is the routine use of the mean of two-line observations in the adjustment, rather than selecting the observation that minimizes loop closure errors. The consequences of this approach are illustrated in Figure 8. To generate the figure, we fixed the geopotential number of the Kauklahti benchmark (based on Saaranen et al., 2021) and recalculated the adjusted geopotential numbers for the rest of the network using our proposed method. We then computed the differences between these adjusted values and the official results.

Figure 8 reveals a clear systematic tilt in the South–North direction, averaging about 4 mgpu per degree of latitude. This suggests that one potential cause of such tilting in levelling networks (see Gerlach and Rummel, 2024) is the systematic bias introduced by averaging two observations rather than selecting the most consistent one. A secondary, though smaller, contributor to the tilt is the use of weights that are inversely proportional to levelling line lengths.

5. Conclusion

The primary objective of this research was to expand on Cvetkov's study (Cvetkov, 2024d) by introducing new weighting strategies and reconfiguring the analyzed levelling network as a free network, with the datum positioned at its mass centre. The results show that the adjustment algorithm presented here reduces the uncertainty of adjusted benchmark geopotential numbers by a factor of 15 to 20. The key step in the algorithm is selecting observations - or their means - that minimise levelling loop closure errors.

The accuracy achieved confirms that precise levelling remains the most reliable method for measuring height differences and investigations. Investigations similar to Izvoltona et al. (Izvoltova et al., 2024) are necessary to evaluate the method. GNSS-based alternatives, while valuable, cannot yet match this level of precision (Apollo et al., 2021; Borowski et al., 2025; Celms et al. 2024; Kurtz et al., 2024). However, precise levelling also has its limitations (Borowski et al., 2025). For this reason, combining terrestrial and satellite-based methods offers the most practical path forward in developing modern vertical reference systems (Borowski et al., 2025; Gerlach and Rummel, 2024).

Leveraging the strengths of both approaches, geoid-based vertical reference frames with uncertainties below 10 mgpu are likely within reach.

Acknowledgements

The authors express gratitude to the Centre for Scientific Research and Design at the UACEG for funding and support for the completed research project under contract No BN315/2025 on the topic "Optimisation of levelling networks by studying the applicability of resampling methods."

References

Apollo, M., Mostowska, J., Maciuk, K., Wengel, Y., Jones, T. E., Cheer, J. M. (2021). Peak-bagging and cartographic misrepresentations: a call to correction. *Current Issues in Tourism*, 24(14), 1970–1975. https://doi.org/10.1080/13683500.2020.1812541

Borowski, L., Banasik, P., Maciuk, K. (2025). Application of GNSS-levelling for updating the base vertical network, *Journal of Applied Geodesy*, DOI: 10.1515/jag-2024-0096.

Celms, A., Grinberga, L., Lidumnieks, T., Luksa, J., Brinkmanis-Brimanis, M. (2024). Assessment of accuracy of GNSS measurement models using base station and radio module solutions. *Baltic J. Modern Computing*, 12(4):454–462. https://doi.org/10.22364/bjmc.2024.12.4.07

Cvetkov, V. (2022). Discrepancies and misclosures in the Third Levelling of Bulgaria, *Annual of the University of Architecture, Civil Engineering and Geodesy, Sofia*, 55(1):195-201.

Cvetkov V. (2023). Two adjustments of the second levelling of Finland by using nonconventional weights, *Journal of Geodetic Science*, 13(1):20220148, https://doi.org/10.1515/jogs-2022-0148.

Cvetkov, V. (2024). A myth concerning the geometric levelling: closing errors and loop circumferences. Proceedings of 24th International Multidisciplinary Scientific GeoConference SGEM 2024, 24(2.1):149-154, DOI: 10.5593/sgem2024/2.1/s09.19.

Cvetkov, V. (2024). Averages, entropy and observations, *Deutsche Internationale Zeitschrift für zeitgenössische Wissenschaft*, 93:52–7.

Cvetkov, V. (2024). Entropy, observations and some standard deviations, *Deutsche Internationale Zeitschrift für zeitgenössische Wissenschaft*, 94:13–8.

Cvetkov, V. (2024). On initial data in adjustments of the geometric levelling networks (on the mean of paired observations). *Journal of Geodetic Science*, 14(1):20220170. DOI: 10.1515/jogs-2022-0170.

Cvetkov, V. (2024). Geometric Levelling Data and Some Systematic Faults in Their Treatment, *Inżynieria Mineralna - Journal of the Polish Mineral Engineering Society*, 2024, 1(1), 293–300. https://doi.org/10.29227/IM-2024-01-32

Cvetkov, V. (2024). *Adjustment of the precise levelling networks*, Monography, Sofia, ISBN 978-619-92816-0-4. (in Bulgarian).

Duchnowski, R., Wyszkowska, P. (2019). Empirical influence functions of different robust estimation methods applied in displacement analysis. *4th Joint International Symposium on Deformation Monitoring (JISDM)*, 15-17 May 2019, Athens, Greece.

Gerlach, C., Rummel R. (2024). Benefit of classical leveling for geoid-based vertical reference frames. *Journal of Geodesy*, 98(64). DOI: 10.1007/s00190-024-01849-y.

Izvoltova, J., Chromcak, J., Bacova, D. (2024). The Effect of Critical Distance in Digital Levelling. *Computation*, 12(6), 111. https://doi.org/10.3390/computation12060111.

Javadi, P. (2017). Detecting the Unstable Points in Deformation Monitoring Geodetic Networks in Analysis Method of Subnetwork. Modern Applied Science, 11(3):61-75, ISSN 1913-1844, E-ISSN 1913-1852, http://dx.doi.org/10.5539/mas.v11n3p61.

Kääriäinen, E. (1966). The Second Levelling of Finland in 1935–1955, *Publications of the Finnish Geodetic Institute No. 61*, Helsinki.

Kurtz, B., Gómez D., Bevis, M. (2024). Characterization of the precision of PPP solutions as a function of latitude and session length, *Journal of Geodetic Science*, 14(1):20220176. DOI: 10.1515/jogs-2022-0176.

Lyszkowicz, A., Leonczyk, M. (2006) The Fourth Precise Levelling Campaign of Poland in 1999–2003, Promoting Land Administration and Good Governance, *5th FIG Regional Conference Accra*, Ghana, 8-11 March 2006.

Saaranen, V., Lehmuskoski, P., Takalo M., Rouhiainen, P. (2021). The Third Precise Levelling of Finland, *FGI Publications No. 161*, Kirkkonummi, https://helda.helsinki.fi/items/921bf91e-e93d-4ccd-9f83-aac063ee014f.

Received December 2, 2025, revised January 12, 2026, accepted February 18, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 202-216
https://doi.org/10.22364/bjmc.2026.14.1.11

Neural Networks and Knowledge Integration: Ontology-Driven Text Encoding for CLIP

Oleksandr HALUSHKA, Viktor SHYNKARENKO

Department of Computer Information Technology, Ukrainian State University of Science and Technologies, Dnipro, Ukraine

oleksandr.halushka01@gmail.com, shinkarenko_vi@ua.fm

ORCID 0009-0005-3447-2676, ORCID 0000-0001-8738-7225

Abstract. This paper presents a method for improving the quality of multimodal "Text-Image" representations by integrating external ontological knowledge into a text query. The proposed approach integrates semantic and associative knowledge from ontologies directly into the text module of the CLIP neural network model through a tree-like syntax structure, similar to the K-BERT method. Adapting the syntax tree for use as an input parameter of the text module involves modifying the original positional encoding mechanism and attention mechanism. Relevant research was conducted on ways to modify the attention mechanism in accordance with the ontological nature of the syntax tree. The results of the experiments showed that the modified model outperforms the basic CLIP in terms of image classification accuracy in a number of object categories. A particularly noticeable increase in quality is observed for specific classes that require additional conceptual knowledge. The article discusses in detail the model architecture, the process of knowledge enrichment, and the impact of the proposed modifications on the model's performance.

Keywords: ontology, neural networks, deep learning, knowledge graph, knowledge engineering, multimodal representations, software, information technologies.

1. Introduction

Multimodal models trained to semantically match information presented in different modalities have achieved significant success in various combinations of applications, such as "text-audio", "text-image", and "text-video". Existing models show serious results in both downscale and upscale tasks. They can generate video from a text description, or vice versa, summarize the content presented in the video, link text, video, and audio content, forming a single semantic context. This is achieved by projecting information presented in different modalities into a single semantic space formed during the model training process. The organization of this semantic space determines how the model builds associative links between fragments of information.

With some simplification, the process of training such models can be characterized as the process of forming their own implicit ontology based on statistical data. The main role of any ontology, whether explicitly or implicitly defined, is to structurally represent knowledge about a certain subject area so that, using it as a source of knowledge, the

interpretation of specific input information is simplified. Then ontology is essentially a generalized representation of some stable patterns or images formed on the basis of individual concepts with associative connections.

Using ontological knowledge at the level of association building, the idea of using explicitly constructed ontologies (ontologies that were constructed by humans) to enhance the implicit internal ontology of the model makes sense. Then a model that receives not only a text query as input, but also some semantic context, will be able to interpret it more accurately.

However, models trained only on statistical relationships between images and text descriptions may lack knowledge about the subject area and miss some important details. Unlike humans, who can rely on broad contextual experience and facts when recognizing an object, neural network models do not have built-in ontological knowledge or modules and methods for acquiring this knowledge. For example, humans know that a "swan" is a bird that can be white or black, lives in pairs, and is larger than a goose. Such knowledge can help distinguish a swan from other birds that look similar. A neural network model forms such knowledge implicitly, statistically, and exclusively from input data, which may not always be effective, especially for specialized subject areas.

One promising direction for improving the quality and accuracy of models capable of zero-shot classification (Wang et al., 2019) is the integration of knowledge from external structured sources, such as ontologies and knowledge graphs. These approaches allow enriching the information presented in one of the modalities with additional semantic context, which aligns the semantic meaning of the information with the generally accepted understanding of the nature of certain concepts.

Based on this approach, this article presents a modified version of the CLIP model (Radford et al., 2021) (hereinafter K-CLIP) that uses external ontological knowledge to enrich text queries. In K-CLIP, the text encoder is modified so that it can accept as an input parameter a syntactic structure similar to text, but with branches inside, which represents a certain concept enriched with relevant ontological triples. The idea is to allow the CLIP model to use explicit knowledge from external sources to improve the quality of its work.

2. Related Works

Modern deep learning models for matching image and text representations are based on the statistical formation of an understanding of the similarity of information presented in different modalities in the process of training on large volumes of image-text pairs.

One of the well-known basic multimodal models is the CLIP (Radford et al., 2021) (Contrastive Image-Language Pretraining) model proposed by OpenAI. This model is trained to compare two modalities: visual and textual. The model was trained on a large-scale corpus of image-text pairs (about 400 million pairs). In these pairs, the text element describes what is depicted in the image. CLIP consists of two main components – text and visual encoders, each of which constructs a representation of the input information of a certain modality into a high-dimensional vector, which represents the "meaning" of what is depicted in the image or described in the text query. Figure 1 shows a diagram of how the model works.

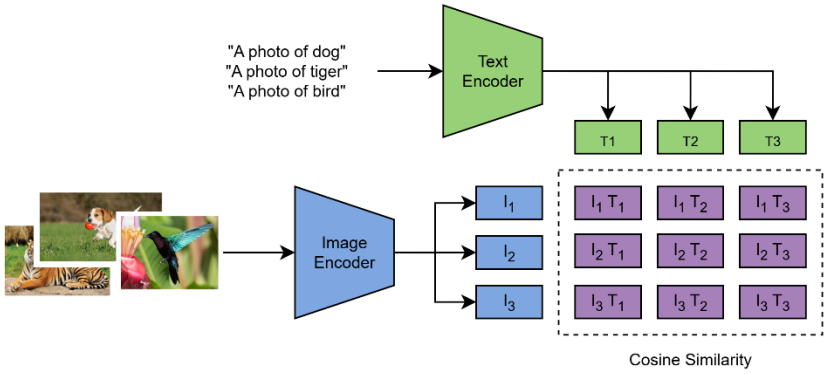


Figure 1. CLIP model workflow

For things that are similar in meaning in the text and in the image, CLIP will construct similar vector representations. Cosine similarity is used as a similarity metric, which compares the direction of vectors relative to each other:

$$Similarity = cos(\theta) = \frac{R_T \cdot R_I}{||R_T||\,||R_I||}, \quad (1)$$

where R_T is the representation of the text query, R_I is the representation of the image, $||R_T||$ and $||R_I||$ are norms of the vectors of the text and visual representations, respectively. The cosine similarity values determine the following properties:

- values close to one indicate semantic correlation between concepts;
- zero values reflect the orthogonality of vectors in multidimensional space, which is interpreted as the semantic independence of concepts;
- negative values close to -1 indicate semantic opposition between concepts.

Using this model, classification can be performed based on similarity metrics. Interestingly, since the model built an implicit ontology during training, which it uses to understand the world, it is capable of zero-shot classification (classification of previously unknown classes that were not represented in the training sample) without additional training.

With its release, CLIP became a landmark work in the field of multimodal representations. The model demonstrated that training on unlabeled correspondences between two modalities can result in universal representations that are independent of a specific modality. A similar approach, ALIGN (Jia et al., 2021), was proposed by Google, where an even larger corpus of image pairs and their text descriptions was used for training. This also led to improved representation quality, which in turn improved results on tasks such as image classification and search.

However, neither CLIP nor ALIGN explicitly use external ontological knowledge. Their internal understanding of the world is formed statistically and therefore does not guarantee the assimilation of specific facts or conceptual hierarchies of concepts.

Early work on training recognition models with semantics in mind showed the advantages of using structured, explicitly specified knowledge. For example, the DeViSE model (Frome et al., 2013) mapped images and words to a common semantic space using

vector representations of words, which allowed objects to be classified even if they had not been encountered during training. This was achieved through the semantic proximity of their names. Another approach is to use a knowledge graph to improve classification (Marino et al., 2016). This approach is based on the use of a graph neural network whose configuration is built on the basis of the topology of a knowledge graph or ontology. This allows the use of knowledge about the relationships between objects in the process of image recognition, which has proven beneficial in multi-class classification.

In the field of natural language processing, the integration of external knowledge has been studied for a long time. The BERT model, which was trained on a large-scale text corpus, does not contain explicit facts. To solve this problem, some modifications of this model were created, such as ERNIE (Zhang et al., 2019), KnowBERT (Peters et al., 2019), and K-BERT (Liu et al., 2020). According to the results, these modified models show better performance on tasks that require specialized knowledge compared to the basic BERT model.

For example, KnowBERT embeds knowledge from WordNet (Fellbaum, 2010) directly into the model parameters, which obviously requires additional training, complicating integration. K-BERT, on the other hand, uses external ontological knowledge as context. K-BERT is most similar to K-CLIP. It differs from other approaches in that it does not require training adapters that translate ontological triples into vector representations. Instead, it enriches the input text with additional ontological knowledge. This process consists of three main stages: recognizing named entities in the text, searching for ontological facts in the knowledge base that correspond to the recognized entities, and embedding the found facts into the sentence structure of the text. Thus, in general, the text retains its basic structure and acts as a kind of skeleton onto which additional information in the form of knowledge is built. However, despite the conceptual similarity in using visibility matrices for knowledge integration, K-BERT and K-CLIP operate in fundamentally different settings: K-BERT modifies BERT, a text-only encoder designed for NLP tasks, and requires task-specific fine-tuning, whereas K-CLIP modifies CLIP, a vision-language model, and operates without additional training. This architectural difference makes direct quantitative comparison between the two approaches not meaningful.

There is also another approach that modifies the CLIP model through prompt tuning using knowledge. The idea is to use the generation of text prompts that describe a certain class of objects based on explicitly specified knowledge (Kan et al., 2023).

A body of previous work points to the potential effectiveness of applying ontological knowledge to improve the performance of multimodal models. The K-CLIP modification continues this work in this direction, architecturally integrating the use of ontological knowledge into the CLIP model for image classification tasks.

3. Methodology

This section describes the architecture of the modification of the original CLIP model to enable the integration of external ontological knowledge. CLIP consists of two encoders: visual (for images) and text (for captions or class names). We leave the visual encoder unchanged, while focusing our modifications on the text encoder.

3.1. Model structure and text enrichment using knowledge

The CLIP text encoder belongs to the family of encoder-only transformers (Singh, 2024). Such transformers accept some data as input and generate their representation in the form of contextually enriched high-dimensional vectors as output. Similar to the BERT model (Devlin et al., 2019), the semantic representation of the entire sequence of tokens as a whole, rather than each token individually, is stored in a special reserved token <CLS>, which is used as a vector representation of the input text. In K-CLIP, the input to the text encoder is not only the concept, but also the tokens added to it that represent knowledge.

The process of knowledge enrichment occurs as follows:

- the knowledge base is searched for ontological facts in which the concept appears as a subject or object;
- the facts found are selected taking into account the degree of semantic depth and quantity;
- a tree-like syntax structure is formed, in which the initial concept with branches acts as the root.

Fig. 2 shows an example of a tree-like syntactic structure built on the basis of an ontological knowledge segment.

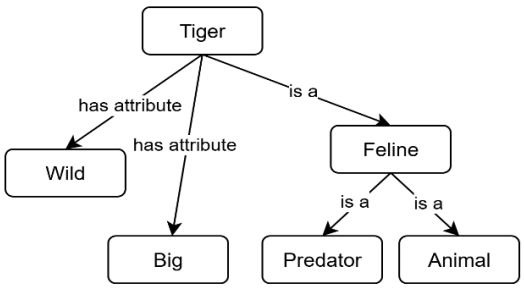


Figure 2a. Ontological segment

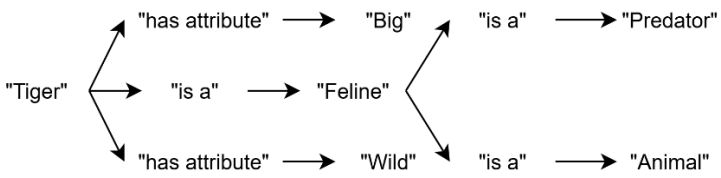


Fig 2b. Tree-like syntax structure based on the segment shown in Figure 2.a

The nature of the knowledge used should, to one degree or another, determine the attributive, semantic, and associative context of the concept.

The structure of the constructed syntactic structure does not coincide with the structure of ordinary text written in natural language. In ordinary texts, each word has its own unique position in the sequence, based on which it can be uniquely identified.

The tree-like structure requires the introduction of relative positioning of tokens in the sequence according to the tree topology. The need for relative positioning arises from the structural difference between linear text and tree-based ontological representations. In natural language, word order is inherently sequential and semantically significant. In contrast, a tree constructed from ontological triples exhibits hierarchical rather than sequential semantics: the order within each triple matters, but different branches representing independent facts have no inherent ordering. Applying standard positional encoding to a flattened tree would incorrectly suggest sequential relationships between topologically independent nodes, distorting the semantic structure that the ontology encodes.

The point is that when transforming a tree into a sequence, each token must be assigned two indexes: actual and relative. The actual index is the position of the token in the entire sequence, which does not affect anything due to the parallel nature of the transformer architecture - it does not matter in what order we submit tokens to the input, but how the positional encoding of tokens is performed. The relative index is the position of the token relative to the initial token, which is determined topologically based on a tree-like structure. The CLIP text encoder uses trained embedding as positional encoding, i.e., by default, the token representation t with index i is appended with positional embedding p with the same index, where i is the actual position of the token in the input sequence.

$$e[i] = x[i] + p[i], \tag{2}$$

where $e[i]$ is the resulting vector of element-wise addition of the components of the token representation x and the corresponding positional embedding p.

Adaptation of the syntactic structure with built-in additional knowledge involves modifying the positional encoding by selecting the corresponding positional embedding using its relative index instead of the actual one. Fig. 3 illustrates both the original CLIP positional encoding mechanism and the proposed modification for tree-structured input.

This approach allows us to approximate the representation of related concepts and the relationships between them at the level of perception by the trained transformer model. Transforming a tree-like structure into a simple sequence destroys the semantic structure, so positional modification helps the model to correctly perceive information without violating its semantic coherence. As discussed above, standard positional encoding assigns sequential positions that do not reflect the tree topology. Consequently, tokens from independent branches appear sequentially adjacent, causing the model to establish spurious semantic connections between unrelated concepts.

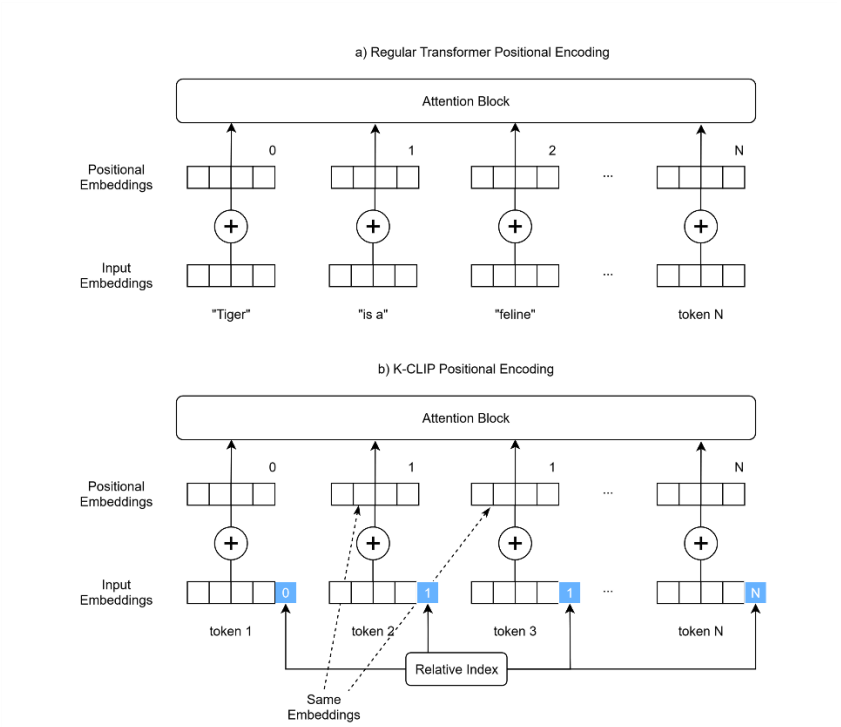


Figure 3. Positional encoding: (a) original CLIP mechanism using actual token positions; (b) K-CLIP modification using relative positions based on tree topology

However, this implementation raises the following problem: different nodes of the tree-like structure with the same relative positional indices will be perceived by the model as being next to each other, which leads to variability in token following. This problem is known as “knowledge noise” and can be described as a situation where the model, using the attention mechanism, establishes a connection between tree nodes that are located in different branches and should not interact with each other in any way. To solve this problem, as well as to control the influence of knowledge, a modification of the attention mechanism is introduced.

3.2. Attention distribution control

The attention mechanism as a concept is a key component of any model built on the transformer architecture. It allows you to form vector representations of tokens in a sequence based on their own context. This means that each token in the sequence correlates to some extent with every other token in that sequence, i.e., the representation of each token distributes its attention among the other tokens in the sequence, including itself. The degree of this distribution allows us to calculate the contribution of each token to the resulting contextual representation of that token.

In the context of K-CLIP, the modification of the attention mechanism consists in limiting the distribution of attention to those tokens that are topologically located in

different branches of the syntactic structure. Each token should only see those tokens that are in the same branch as it is relative to the main concept. The limitation of attention is that topologically distant tokens do not affect the calculation of each other's contextual representation, which prevents the distortion of the semantic meaning of the entire information as a whole. The attention distribution control scheme is illustrated in Fig. 4.

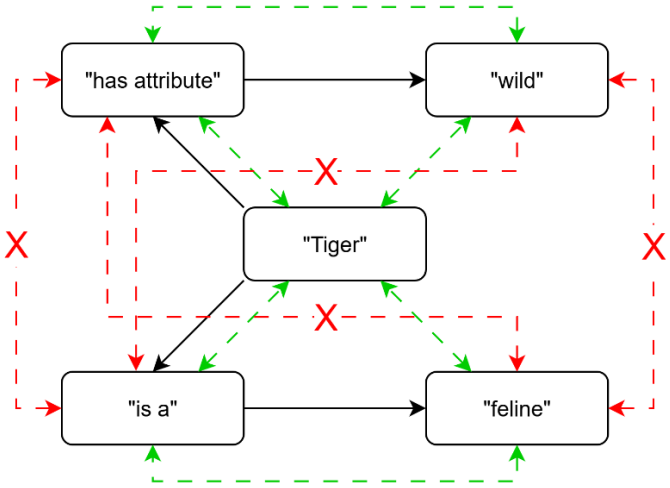


Figure 4. Example of attention distribution control

The formal implementation of the modification was done by constructing a binary visibility matrix VM, in which each element $VM_{i,j} = 1$, if the tokens i and j are in the same branch, and $VM_{i,j} = 0$ if they are in different branches. In terms of size, this matrix coincides with the attention matrix, so to put it into action, each element of the attention matrix A must be multiplied by the corresponding element of the matrix VM:

$$MA = A \circ VM, \tag{3}$$

where MA is the modified attention matrix formed by element-by-element multiplication of the corresponding elements of the matrices (operation $\circ$), A is the original attention matrix, and VM is the visibility matrix.

The visibility matrix dimensions are bounded by the CLIP text encoder's context window of 77 tokens, resulting in a maximum matrix size of 77 × 77. After accounting for special tokens (<CLS>, <SEP>) and the root concept, approximately 72 tokens remain available for ontological knowledge. Given that a typical ontological triple requires 6–10 tokens after tokenization (subject, predicate, and object with potential subword splitting), the practical ontology size is constrained to 7–12 triples per query. In our experiments, we used 10–12 facts per concept, which approaches this architectural limit while maintaining representational quality. This constraint does not fundamentally limit the number of relationships that can be encoded, but rather requires careful selection of the most

semantically relevant triples for each concept.The algorithm for constructing the visibility matrix is presented below:

Algorithm 1. Visibility Matrix Construction

```
Input: tree T with root r, token sequence {t₁, ..., tₙ}
Output: visibility matrix VM ∈ {0,1}^(nxn)

1: for each token tᵢ do
2:   path[i] ← ancestors(tᵢ, r) ∪ {tᵢ}  // path from root to tᵢ, including tᵢ
3: end for
4: for i = 1 to n do
5:   for j = 1 to n do
6:     if tᵢ ∈ path[j] OR tⱼ ∈ path[i] then
7:       VM[i,j] ← 1  // tokens are in the same branch
8:     else
9:       VM[i,j] ← 0  // tokens in different branches
10:    end if
11:  end for
12: end for
13: return VM
```

An example of a generated visibility matrix is shown in Fig. 5.

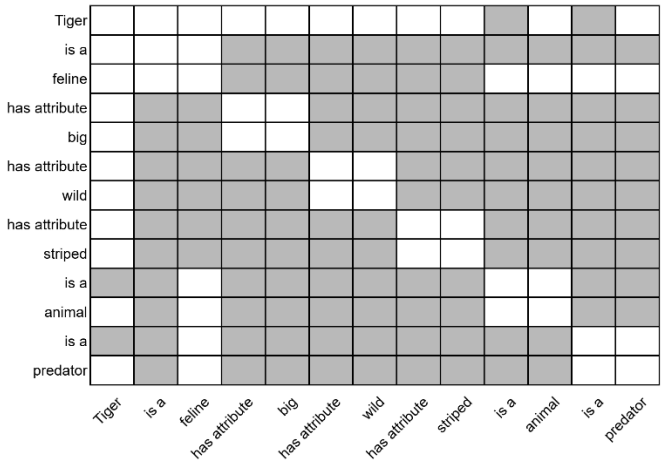


Figure 5. Visibility matrix

This approach to controlling the distribution and concentration of attention ensures that the main root concept will play a decisive role in the resulting representation, and that the additional knowledge with which this concept has been enriched will not interact with each other at the level of attention. This follows from the very idea of enriching some input data with external knowledge. Knowledge should semantically “guide” the representation of the concept closer to its semantic meaning, rather than obscuring the initial representation.

4. Experiments

To evaluate the effectiveness of the approach, a series of experiments was conducted on image classification tasks for different categories and classes of objects. The CLIP model with a visual encoder based on ViT-B/32 (Awofeso, 2024) and a text encoder with a vector dimension of 512 (the original model) was used as a basis. The model was initialized with the trained parameters of the CLIP model without additional training. Small knowledge graphs with approximately 10-12 facts each were used as ontological knowledge. The triplets of these graphs describe primarily visual attributes, as well as hierarchical and semantic components of the concept.

The experiments were conducted in three stages:

- testing the variability of the attention mechanism modification for different blocks in order to determine how much the modification at different levels of abstraction affects the quality of the model's performance;
- testing the model for the effectiveness of object classification using additional knowledge as a source of additional clarification and awareness in order to establish the effectiveness of the influence of external knowledge on the quality of classification;
- testing the model's ability to perform fine-grained discrimination between visually similar object classes (different dog breeds) by comparing three encoding strategies: generic concept baseline, K-CLIP with breed-specific ontological attributes, and direct breed naming.

The first stage is to check how modifying attention on different sequential blocks of the Multi Head Attention transformer component affects the result. The CLIP model, like any other model based on transformer architecture, includes a Multi Head Attention component, which is represented as a sequence of interconnected attention blocks. A typical model usually has 12 blocks. The sequential connection of blocks allows attention to be formed at different levels of abstraction of information perception. The first blocks establish connections between atomic fragments, while the latter direct attention between more abstract concepts represented in the input token sequence. Tests were conducted on different configurations in which some blocks were modified while others were not.

In this experiment, we investigate the effect of applying our modifications at different layers of the transformer. The K-CLIP approach involves two components: (1) relative positional encoding, which is applied uniformly to all input tokens, and (2) attention masking via the visibility matrix, which can be selectively enabled at specific transformer blocks. In these experiments, we vary which blocks receive the attention masking while keeping relative positional encoding constant. The tested configurations are summarized in Table 1.

Table 1. Attention masking configurations

Configuration	Modified Blocks	Description
First only	Block 1	Masking at the lowest abstraction level
Middle only	Block N/2	Masking at intermediate abstraction level
Last only	Block N	Masking at the highest abstraction level
Both ends	Blocks 1 and N	Combined low and high level masking
All blocks	Blocks 1-N	Full masking throughout all layers
None	-	Baseline

Such variability in configurations helps to reveal the dependence of the effect of attention modification on the level of abstraction of representation. This provides an understanding of in which cases and under which configurations attention modification in “manual mode” can take place, and how this affects the quality of the model's performance. The main dataset was used as a test data set, the data of which was divided into three categories: animals, rare birds, and rare things. The animals category included the following object classes: cat, dog, eagle, elephant, ostrich, rabbit, tiger; the rare birds category included the following object classes: helmeted hornbill, imperial woodpecker, kakapo, purple-throated carib, california condor; the rare things category included the following object classes: bathynomus giganteus, grandidierite, hallucigenia sparsa, welwitschia mirabilis. For the experiments, 10 images were selected from each class (160 in total). Fig. 6 shows some examples of images from the test data set.

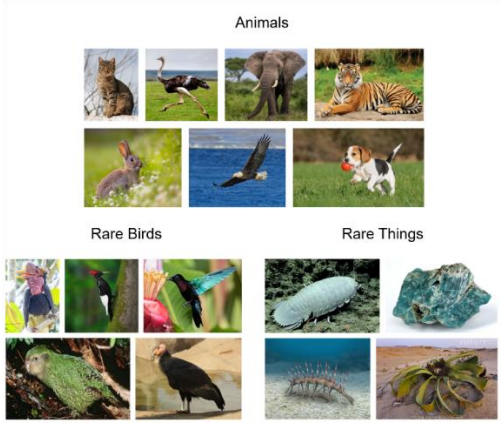


Figure 6. Object categories

The dataset from the previous stage was used for the second stage of the experiments. The quality of classification was evaluated by the relative deviation of the cosine similarity value of the modified model from the value obtained from the original model. Given that cosine similarity is used, a positive deviation indicates an improvement in quality, and a negative deviation indicates the opposite, since when comparing information of different modalities with the same semantic meaning, we expect the ideal result to be close to unity.

During the third stage, the model was tested for its ability to distinguish between different dog breeds. Five dog breeds were selected: German Shepherd, Pekingese, Saint Bernard, Welsh Corgi, and Yorkshire Terrier. Ontologies were formed for each breed, each of which includes 10-12 facts that focus more on the original visual attributes characteristic of a particular breed. Twenty images of dogs were also selected for each breed (100 in total).

The quality of how well the model distinguishes one breed from another was assessed by comparing cosine similarity values across three encoding strategies: using only the generic base concept "dog" (same for all breeds), using K-CLIP with breed-specific ontological knowledge, and using the direct breed name. This experimental design allows evaluating the contribution of ontological knowledge to fine-grained discrimination.

5. Results

5.1. Variative modification of attention

The results of the first stage of experiments show that the trained model significantly degrades in quality when modifying the initial blocks responsible for more specific and atomic aspects of the input information. Modification of the first block in all cases and combinations gives a negative result of approximately -47% relative to the original model without the use of external knowledge. At the same time, modifying attention only on the last block gives a positive effect and an increase in the quality of the model's performance by 3.16% relative to the original model. Fig. 7 shows the results of the modified model with different configurations of block modifications. The indicator is the increase in representation quality relative to the original CLIP model. Negative values indicate a deterioration in performance, while positive values indicate an improvement in the quality of representation construction.

The results show that manipulating the attention mechanism is very sensitive and can only be applied to the final blocks when it is necessary to slightly direct attention in the right direction without destroying the basic structure of the attention calculation process and the connections between the parameters of the trained model. The sensitivity stems from how pre-trained transformers organize information processing. Early attention blocks have learned to capture local syntactic patterns and basic token dependencies; disrupting these patterns invalidates the representations that all subsequent layers expect as input. Final blocks, however, refine already-coherent semantic representations, making them more tolerant to controlled modifications that guide attention toward specific semantic aspects without breaking the model's internal consistency. Importantly, this approach allows engineering semantic relationships without requiring explicit knowledge of which internal vectors correspond to which concepts. Instead of manipulating representations directly, we control relationships structurally through the visibility matrix, leveraging the transformer's own attention mechanism to respect ontological structure.

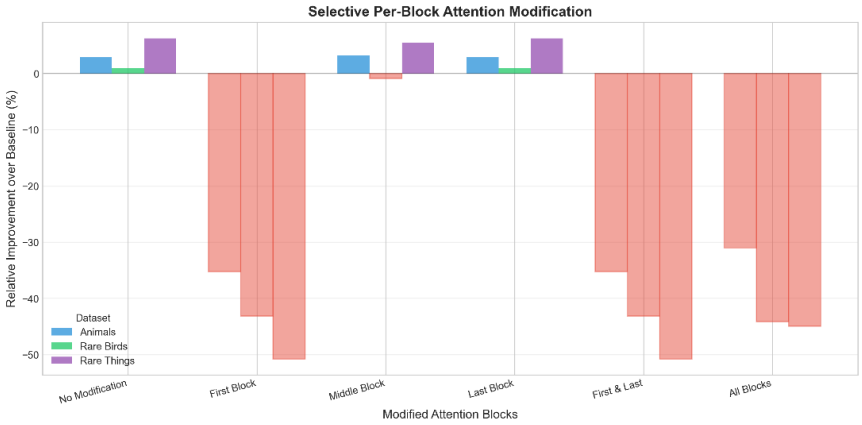


Figure 7. Modification of attention in different configurations

5.2. Classification

Tests of the model's ability to classify and compare its results with the initial unmodified model show a positive increase in efficiency, especially for those classes of objects that are rarely found in the information space and have names that are not prone to strong associative reflections. Thus, the rare things category showed an average increase of 6.22%, which is the best indicator compared to the other two categories, animals and rare birds, with results of 2.5% and 1.3%, respectively. It should be noted that generating text descriptions based on ontology using a third-party LLM gives better results than embedding knowledge. The average increase in efficiency compared to the unmodified model for the categories animals and rare things was 6.5% and 6.75%, respectively. These results can be explained by the fact that the original model was trained specifically on text descriptions, so it perceives plain text better than adapted ontological triples. The results of the experiment are presented in Table 2.

Table 2. Classification of different categories of objects

Category	S_1	S_2	S_3	$S_2 \rightarrow S_1$, %	$S_3 \rightarrow S_1$, %
Animals	0.28	0.29	0.30	2.50	6.50
Rare birds	0.33	0.33	0.33	1.30	0.78
Rare things	0.30	0.32	0.32	6.22	6.75
Average	0.30	0.31	0.31	3.15	4.88

In the table: S_1 is cosine similarity between the image and the object name calculated by the original model, S_2 is the modified model using knowledge embedding, S_3 is the similarity between the image and the synthetically generated description based on ontological triples.

5.3. Discrimination of subtle object categories

Assessing the model's ability to distinguish between conceptually and semantically similar classes of objects demonstrates the value of additional knowledge in fine-grained discrimination tasks. We compare three encoding strategies: S_1 uses only the generic concept "dog" as a baseline, S_2 uses K-CLIP with breed-specific ontological attributes, and S_3 uses the direct breed name. This progression allows us to isolate the contribution of structured ontological knowledge ($S_2 \rightarrow S_1$) from the upper bound achievable with explicit naming ($S_3 \rightarrow S_2$). The results are presented in Table 3.

Table 3. Distinguishing between similar classes of objects

Breed	S_1	S_2	S_3	$S_2 \rightarrow S_1$, %	$S_3 \rightarrow S_2$, %
German Shepherd	0.27	0.30	0.31	+12.1	+5.8
Pekingese	0.26	0.30	0.35	+13.4	+15.4
Saint Bernard	0.26	0.27	0.33	+4.8	+20.1
Welsh Corgi	0.27	0.29	0.32	+6.7	+11.2
Yorkshire Terrier	0.25	0.28	0.34	+10.8	+19.3
Total	0.26	0.29	0.33	+9.6	+14.4

In the table S_1 is the cosine similarity using only the base concept "dog" (same embedding for all breeds), S_2 is the similarity using K-CLIP with breed-specific ontological attributes, S_3 is the similarity using the direct breed name. The columns $S_2 \rightarrow S_1$ and $S_3 \rightarrow S_2$ show the relative improvement in similarity.

6. Conclusions

This paper presented the K-CLIP model, a model that uses external ontological knowledge as a source for semantic enrichment of input text information. The approach is based on the integration of ontological triples into a text encoder by introducing a mechanism for topological positioning of tokens and controlling the distribution of attention flow in a trained model built on the basis of transformer architecture. This solution allows the model to take into account semantic relationships based on the ontological facts provided to it when comparing images with textual information, which brings its behavior closer to the human way of comparing multimodal information: when a person sees an object, they also use their previous experience as a source of knowledge, which helps them identify the object more accurately and confidently.

The experiments confirmed the practical promise of this approach, as the modified model outperformed the original in a series of classification tests, especially where it was necessary to distinguish subtle semantic differences, such as in the experiment on classifying different breeds of dogs, where in each syntactic tree structure the concept of "dog" served as the initial concept, and the differences were only in the additional knowledge that was attached to this concept depending on the specific breed.

The study opens up opportunities for further work in the direction of using formal knowledge as a way to improve the efficiency of neural network models. In the future, it is planned to improve methods of using structured knowledge to increase the interpretability of deep learning models and their transition from simple pattern recognition to semantic understanding of context.

References

Awofeso, Z. (2024). An explanation of the Vision Transformer (ViT) paper. Medium. https://medium.com/codex/an-explanation-of-the-vision-transformer-vit-paper-8cdd399741aa

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, volume 1 (long and short papers) (pp. 4171-4186).

Fellbaum, C. (2010). WordNet. In Theory and applications of ontology: computer applications (pp. 231-243). Dordrecht: Springer Netherlands.

Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Ranzato, M. A., Mikolov, T. (2013). Devise: A deep visual-semantic embedding model. *Advances in neural information processing systems*, 26.

Jia, C., Yang, Y., Xia, Y., Chen, Y. T., Parekh, Z., Pham, H., ... Duerig, T. (2021). Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning* (pp. 4904-4916). PMLR.

Kan, B., Wang, T., Lu, W., Zhen, X., Guan, W., Zheng, F. (2023). Knowledge-aware prompt tuning for generalizable vision-language models.

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P. (2020). K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 03, pp. 2901-2908).

Marino, K., Salakhutdinov, R., Gupta, A. (2016). The more you know: Using knowledge graphs for image classification. arXiv preprint arXiv:1612.04844.

Peters, M. E., Neumann, M., Logan IV, R. L., Schwartz, R., Joshi, V., Singh, S., Smith, N. A. (2019). Knowledge enhanced contextual word representations. arXiv preprint arXiv:1909.04164.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.

Singh, R. (2024). Types of Transformer Model. Medium. https://medium.com/@RobuRishabh/types-of-transformer-model-1b52381fa719

Wang, W., Zheng, V. W., Yu, H., Miao, C. (2019). A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology* (TIST), 10(2), 1-37.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q. (2019). ERNIE: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129.

Received October 27, 2025, revised December 18, 2025, accepted February 20, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 217-237
https://doi.org/10.22364/bjmc.2026.14.1.12

Exploring the Connection Between Personality Dimensions and Learning Strategies in Mathematics

Aurelija KASPARAVIČIŪTĖ, Svetlana DANILENKO,
Rūta SIMANAVIČIENĖ

Vilnius Gediminas Technical University, Sauletekio ave. 11, LT-10223 Vilnius, Lithuania

aurelija.kasparaviciute@vilniustech.lt,
svetlana.danilenko@vilniustech.lt,
ruta.simanaviciene@vilniustech.lt

ORCID 0000-0001-8261-314X, ORCID 0000-0001-8731-2522,
ORCID 0000-0002-1614-9470

Abstract. Personality traits are indicative of consistent patterns in thoughts, emotions, and behaviours, and they consequently impact students' academic performance. Numerous scientific investigations have explored the determinants of student learning success, with some emphasizing the significance of considering both student personality traits and generational cohorts. This study seeks to examine the associations between the Big Five personality dimensions and the learning strategies employed in mathematics lectures. A random sample of first- and second-year informatics engineering science students from Vilnius Gediminas Technical University (VILNIUS TECH) was selected for this research. Data were gathered through an anonymous survey utilizing a questionnaire developed by the authors. Non-parametric statistical tests were employed to analyse the data. The findings revealed that the sample predominantly consisted of students exhibiting the personality dimensions of agreeableness and conscientiousness. Furthermore, students' selection of learning strategies was generally independent of their personality dimensions. Nonetheless, the study underscored students' preferences for diverse learning strategies, irrespective of their personality traits.

Keywords: Big Five personality dimensions, academic success, mathematics learning strategies, Generation Z.

Introduction

Hendrastomo and Januarti (2023) characterize education as a dynamic concept, asserting that "In the world of education, there is nothing static; the curriculum is ever-evolving, knowledge continually expands, and the participants, both teachers and students, are in a constant state of transformation. <…> Education engenders culture and order, akin to a domestication process. It is also a process of cultivating various human potentials, including academic, relational, talent, physical, and artistic abilities." In contemporary times, education must adapt to address the needs of the new generation, as its role is to enhance, develop behavior, thinking, and action, and to motivate adaptation to the evolving times (Hendrastomo and Januarti, 2023). Numerous studies in educational

research have concentrated on developing more effective teaching and learning strategies (Biwer et al., 2020; Kumari et al., 2023; Munna and Kalam, 2021; Cacula 2024; Pliuskuvienė et al., 2024). Teaching methods encompass a variety of strategies and techniques employed by educators to engage students and improve learning outcomes. Sugano and Mamolo (2021) identify several teaching methodologies, including combination learning, cooperative learning, inquiry-based learning, individualized instruction, manipulative and multiple representations, multicultural education, problem-based learning, project-based learning, and technology-based instruction. In this study, learning strategies are defined as sets of tools (e.g., textbooks, lecture videos, interactive knowledge tests, etc.) and activities (e.g., individual or group consultations, group work, lecture work, homework, etc.) that constitute the learning process, from which the learner makes selections. Similarly, Juste and López (2010) describe learning strategies as approaches that incorporate various steps or sequences of actions that learners choose to enhance their learning experience, address challenges, and meet academic requirements (Juste and López, 2010). Bligh and Cameron (2000) argue that traditional teaching and learning (e.g. teacher-centric, classroom-based, focused on lectures and textbooks methods; teacher has full control over the learning environment, decides the teaching methodology; student learn through repetition and memorization, there is little or no scope for critical thinking) methods are ineffective due to their reliance on economic evaluation and their tendency to foster passive learning. Similarly, Kumari (2023) and colleagues assert in their research paper that "the traditional method of teaching and learning has been in existence for centuries, with little changes over time. However, this approach is no longer effective in the twenty-first century." Consequently, Kumari et al. (2023) aimed to ensure that educational topics are taught in a manner tailored to the needs of individual learners.

Significant attention has been directed towards scientific papers that explore more effective teaching methods and learning strategies, with the goal of identifying the challenges faced in educating Generation Z (Gen Z) and enhancing teaching-learning strategies (Cilliers, 2022). Currently, university students born between 1997 and 2012 are identified as Gen Z. Hendrastomo and Januarti (2023) concluded that Gen Z is distinguished by its unique characteristics. This generation is characterised by a complex interplay of universal values and self-centeredness, which influences their communication behaviour and unique approach to the environment. Furthermore, feedback on achieved results and the integration of the latest technologies in daily activities are crucial for Gen Z (Dolot, 2018). Gen Z students are more self-confident, technologically adept, and focused on personal development (Hernandez-De-Menendez et al., 2008). In the 21st century, there has been considerable discourse on generational differences, with significant emphasis on developing teaching strategies for Gen Z students to address the challenges encountered in educating this cohort (Cilliers, 2022).

Extensive research has demonstrated that a student's academic success is linked to their personality traits. To enhance the accessibility and sustainability of higher education, the relationship between individualized teaching methods and students' academic performance should be examined (Abe, 2020). Lyons et al. (2017) investigated the Flipped Classroom method and observed that student personality traits are a primary factor in academic success. The Flipped classroom method reverses traditional teaching by engaging students to learn new content outside of class and using class time only for active learning, discussions, problem-solving, and collaborative projects with teacher guidance.

Durak (2023) posits that personality traits significantly influence students' academic outcomes. Feher and Vernon (2021) examined five major personality dimensions that can categorize a wide range of personality traits. These dimensions, known as the Big Five, include Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to experience. Accordingly, it can be concluded that considering the personality dimensions of Gen Z may enhance the effectiveness of the study process.

This study aims to investigate whether different learning strategies employed in VilniusTech University mathematics lectures are equally effective for Gen Z students with varying personality dimensions.

1. Related works

As Kumari et al. (2023) observe, learning outcomes in the twenty-first century are frequently evaluated through quantitative indicators such as graduation rates. While such metrics have financial and institutional relevance, they do not necessarily align with broader educational objectives or students' career aspirations. Traditional teaching and learning approaches have been shown to be increasingly ineffective in contemporary educational contexts, as they often fail to prioritise critical thinking, undervalue the development of practical skills, and inadequately respond to the evolving demands of the labour market. To enhance educational effectiveness, there is a growing consensus that instruction must be increasingly individualised and oriented towards the development of competencies essential for the modern era (Kumari et al., 2023). Within higher education, Voskoglou (2019) highlights that traditional teaching methods; particularly in mathematics remain prevalent, sometimes supplemented by challenging questions and mathematical discourse aimed at fostering conceptual understanding and critical analysis. Through a classroom based experimental comparison of traditional and alternative instructional approaches at the university level, Voskoglou (2019) demonstrates the potential value of such pedagogical innovations, while also emphasising the need for further empirical research before definitive conclusions can be drawn. Complementing this pedagogical perspective, Anthony (2000) identifies students' intrinsic motivation as a primary determinant of success in mathematics learning, while Afshar et al. (2019) demonstrate that both lecturers and institutional leadership play a significant role in shaping student motivation in higher education contexts.

An expanding body of research further indicates that individual differences; particularly personality traits are critical predictors of academic success (Durak, 2023; Opoku et al., 2023). Collectively, these findings suggest that improving the effectiveness of mathematics education requires not only refined teaching strategies but also a deeper understanding of learners themselves. Consequently, identifying which learning approaches are most effective necessitates a fundamental question: "who are contemporary university students, and what challenges do they face?".

As noted earlier, the majority of today's university students belong to Gen Z, typically defined as individuals born between 1997 and 2012 (Marin and White, 2023). Hendrastomo and Januarti (2023) argue that educators' efforts to understand the defining characteristics of Gen Z are among the most important factors contributing to effective learning. Gen Z students are described as pragmatic, self-aware, and inclined towards individualised learning; they recognise education as a personal investment, have grown up

within a globalised and rapidly changing society, and often display confidence, competitiveness, curiosity, independence, and a preference for autonomy (Hendrastomo and Januarti, 2023). Lopez and Abadiano (2023) propose a technological, motivational, learning framework to better align instructional practices with the learning preferences of Gen Z. This framework emphasises e-learning, interactive pedagogy, career and financial literacy, interpersonal relationships, healthy competition, multimodal instruction, and the application and contextualisation of concepts. Gen Z students tend to engage more deeply with interactive digital environments and place greater value on applied understanding than on rote memorisation. Given the elevated levels of stress experienced by this cohort, openness to discussions around mental health and social equity is also particularly salient. Moreover, education and career advancement are highly prioritised, underscoring the importance of financial literacy and career guidance (Twenge, 2017). Despite their immersion in digital environments, evidence suggests that Gen Z students often lack the skills required to use digital technologies effectively for learning purposes and may prefer individualised tasks over collaborative work (Marin and White, 2023). Research further indicates that incorporating real-world problems; particularly those arising from social contexts alongside practical tasks that promote reasoning and problem-solving can significantly enhance achievement in mathematics. Lecturers who account for the preferences and characteristics of Gen Z are therefore better positioned to design learning experiences that resonate with students and foster deeper conceptual understanding.

In parallel, substantial research has established strong associations between personality traits and academic performance (Durak, 2023; Çağataylı and Çelebi, 2022; Vedel and Poropat, 2017). Personality traits are commonly defined as relatively stable patterns of thought, emotion, and behaviour that distinguish individuals and can be used to explain and predict outcomes such as academic achievement (Anglim and O'Connor, 2019). While there is ongoing debate regarding the most appropriate psychological framework for categorising personality, the Five Factor Model (FFM), or Big Five model, remains the most widely adopted approach (Feher and Vernon, 2021). Originally proposed by Goldberg (1981, 1990) and further validated by McCrae and Costa (1987), the Big Five model encompasses neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. These dimensions capture emotional stability, sociability, intellectual curiosity, interpersonal orientation, and goal directed behaviour, respectively (McCredie and Kurtz, 2020). Research consistently demonstrates that conscientiousness is a particularly strong predictor of academic achievement, including in mathematics, as it is associated with effective study habits, persistence, and motivation (Muhid et al., 2021). Conversely, elevated levels of anxiety linked to neuroticism can impair the cognitive resources required for mathematical problem-solving, thereby limiting academic success (Koner and Mazumder, 2024).

Recent studies further suggest that predicting academic success solely on the basis of prior grades or entrance examination scores is insufficient (Çağataylı and Çelebi, 2022). Instead, incorporating personality dimensions provides a more nuanced understanding of learning outcomes. Wakefield and Tyler (2023), for example, demonstrate that increased interactivity in theoretical lectures significantly enhances student performance compared with passive learning formats such as recorded videos.

In summary, existing research highlights a substantive relationship among personality traits, instructional design, and student learning outcomes. While considerable progress has been made in identifying effective pedagogical strategies, further investigation is

required to deepen understanding of how individual differences interact with teaching practices. Addressing these factors remains central to improving both the quality and success of higher education.

2. Methodology

This article analyses which learning strategies are more acceptable to Gen Z students based on Big Five personality dimensions (agreeableness, conscientiousness, neuroticism, extraversion and openness). In this study the following learning strategies are included: material provided by the lecturer (e.g. textbooks, lecture videos, interactive knowledge tests, etc.); organizational structure of the lecture (e.g. student are constantly involved in the discussion, on issues related to the course, solving tasks in student groups with the possibility to consult with the lecture, etc.); organizational structure of the study process (e.g. to know in advance assessment system, list of all topics, strictly follow dates of assessments, class attendance, etc.); communication style of the lecturer (e.g. lecturer is charismatic, agreeable, restrained, etc.); student personal learning process (e.g. wants to gain study knowledge, to be one of the best, can set down his own goals, loves to work in groups, etc.), that make up the learning process.

In order to assess the effectiveness of the learning process applied in mathematics lectures, a qualitative study was conducted. This study focused on the relationship between learning tools, activities and student personality dimensions. Initially, a questionnaire (see, WEB, a, Questionnaire) was developed for the first and second-year students of Informatics Engineering at VilniusTech who are enrolled in mathematics courses. The questionnaire was performed at the end of the studies and presented to students in their national language, i.e. Lithuanian. It consists of eight parts, each containing between 10 and 25 questions:

I. A questionnaire developed to assess the five personality dimensions (Bunevičius, 2005).
II. The use of materials presented in the MOODLE environment.
III. The benefits of organisational forms during lectures.
IV. The influence of course organisational structure on the learning process.
V. The impact of the lecturer's communication style.
VI. Factors that may influence personal students' learning process.
VII. Students' ability to plan time and organise their work.
VIII. Demographic indicators.

The first part of the questionnaire consists of 25 questions and is designed to identify personality dimensions (see Part I in WEB, a, Questionnaire). The approach for evaluating personality dimensions is grounded in the framework established by Goldberg (1999). According to this methodology, Bunevičius (2005) developed a questionnaire that uses 25 pairs of adjectives. In each pair of adjectives, the respondents must indicate how strongly one of the adjectives applies to them on a scale from 1 to 7. In Table 1, the labels of the adjectives describing each personality dimension are indicated. We want to remark, that abbreviations of traits are listed in questionnaire (see Part I in WEB, a, Questionnaire).

Table 1. Assessment of the Big Five personality dimensions (Bunevičius, 2005)

Personality dimension	Notation
Extraversion	A, F, K, P, U. Reversed: A, K, U
Conscientiousness	D, I, N, S, X. Reversed: D, N, X
Agreeableness	E, J, O, T, Y. Reversed: J, T
Neuroticism	B, G, L, Q, V. Reversed: G, V
Openness	C, H, M, R, W. Reversed: C, M, W

Each personality dimension in Table 1 is evaluated separately by adding up the scores of the responses of different combinations of pairs of traits. For each personality dimension, the total arithmetic sum of the answers of the corresponding pairs of traits is added separately, except for those marked "Reversed", which are evaluated inversely: "7" corresponds to "1"; "6" - "2"; "5" - "3"; "1" - "7"; "2" - "6"; "3" - "5"; rating "4" is not changed. The dominant personality dimension is the one with the highest arithmetic sum of points.

To determine the factors that impact learning mathematics and how they are linked to students' personality dimensions, six parts of questions starting from Part II, were formulated in the questionnaire (see Parts II-VII in WEB, a, Questionnaire). These questions were developed by the authors based on a review of the scientific literature (Anthony, 2000; Baruth and Cohen, 2023; Çağataylı and Çelebi, 2022; Durak, 2023; Jensen, 2015; Komarraju and Karau, 2005; Opoku et al., 2023; Rajapakshe, 2017; Voskoglou, 2019) as well as the authors' personal teaching experiences. Additionally, the questions from the eighth part of the questionnaire are intended to determine the demographic characteristics of the sample. In more detail, Part II contains questions related to the use of material provided by the lecturer in the MOODLE environment (see Part II in WEB, a Questionnaire, or Table 2). These questions are assigned to assess how often students use the lecturers' prepared course learning tools. Part III (see Part III in WEB, a Questionnaire, or Table 3) consists of questions related to the lecture's organisational structure. These questions are designed to evaluate students' choices of lecture structure, taking into account their own level of information absorption. Part IV (see Part IV in WEB, a Questionnaire, or Table 4) addresses the influence of course organisational structure on the learning process. These questions are assigned to determine how students evaluate the influence of the organisational study process, taking into account students' own learning process. Part V (see Part V in WEB, a Questionnaire, or Table 5) is intended for questions connected to the impact of the lecturer's communication style. These questions are designed to determine how students evaluate the relationship between the lecturer's communication style and the quality of student education. Part VI (see Part VI in WEB, a Questionnaire, or Table 6) contains questions referring to factors that may influence students' personal study learning process. Part VII (see Part VII in WEB, a Questionnaire, or Table 7) involves questions related to students' ability to plan time and organise their own work. Part VIII (see Part VIII in WEB, a Questionnaire) is composed of questions related to demographic indicators, such as gender, age, place of secondary education, study program, etc.

Let us remark that in this paper, the activities and tools selected by the lecturer are considered effective if they are acceptable to the majority of students. The effectiveness of learning strategies is associated with a higher frequency.

The majority of variables analysed in this study are qualitative and measured on an ordinal scale. To assess statistically significant differences between variables, non-parametric tests were applied: for dependent samples, the Friedman test and the Wilcoxon Signed Ranks Test were used; for independent samples, the Kruskal-Wallis test and the Mann-Whitney U Test were employed. For qualitative variables, the dependence between categorical variables was examined using the Chi-Square Test. Differences were considered statistically significant when the p-value (p) < 0.05. The research data were analysed using IBM SPSS Statistics, version 26.

3. Case study

The described methodology was used for a specific case analysis. The population of the study consisted of 108 students from the 1st and 2nd semesters of the Informatics Engineering study programs at VILNIUS TECH university, who studied mathematics courses in their first year. The aim is to identify the factors that impact learning mathematics courses and how they are linked to students' personality dimensions. In the future, this study will be expanded by including students from different study programs.

Based on the research data (see Part VIII in WEB, a, Questionnaire) it can be observed that more male students choose to study informatics engineering. In this study, male students made up 73.63% of the respondents. The majority of the research sample consisted of Software Engineering students (54.44%), followed by Multimedia Design students (27.78%), Information Systems students (13.33%), and Information Technology program students (4.44%). In addition, the majority of respondents (97.83%) studied in a state-funded place, and most of them (70.65%) studied in the desired program. According to VILNIUS TECH university's Admissions and Information Center, most of our students belong to Gen Z (born from 1995 until 2003).

According to the survey results, the vast majority of students participating in the study belonged to the agreeableness (39.13%) and conscientiousness (36.96%) dimensions (see Figure 1). Some respondents were assigned to several dimensions (14.81%). Thus, due to statistical analysis requirements, the further study examined only the data of agreeable, conscious and extroverted students.

Further analysis was performed on data obtained from Parts II-VII (see Parts II-VII, in WEB, a, Questionnaire). Firstly, the aim is to assess whether statistically significant differences between personality dimensions exist or not, considering student preferences for different learning strategies. Secondly, if a significant difference between personality dimensions exists for a given question, it is examined between which personality dimensions and what assessments students made on that question. In addition, box plot diagrams were used to determine which learning strategies students (regardless of personality dimension) prefer.

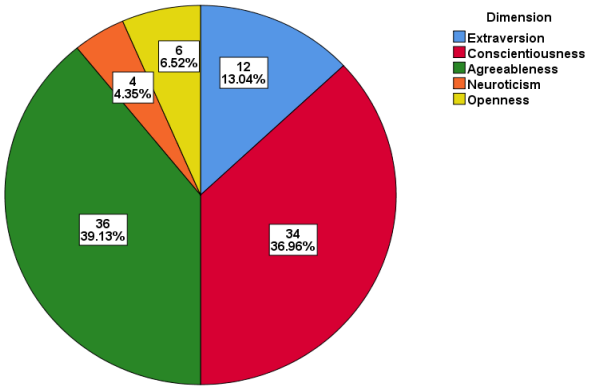


Figure 1. Distribution of students according to personality dimensions

Firstly, the analysis aimed to determine how often students use the material provided by the teacher in the MOODLE environment, depending on its form. To assess whether student preferences for different forms of course materials (see Part II, Question I, in WEB, a, Questionnaire) depend on the student personality dimension, the Kruskal-Wallis test was applied. The study showed that the choice of students in groups of different personality dimensions did not statistically significantly differ. It was concluded by the results of the Kruskal-Wallis test, since for all cases $p > 0.05$ (Table 2). Thus, it means that the choices of course material forms for students with different personality dimensions are similar.

Table 2. Statistical significance difference between personality dimensions' groups, considering students' preferences for different forms of course materials.

Rate how often you use the material provided by the teacher in the MOODLE environment, taking into account its form (0 - I do not use it at all. From 1 - I hardly use it, to 7 - I use it very often)	**p, df =2**
Lecture material (outlines or slides) (q.2.I.1)	0.990
Lecture videos (q.2.I.2)	0.439
Additional lecture material (q.2.I.3)	0.920
Literature (not scientific papers) (q.2.I.4)	0.745
Scientific literature (q.2.I.5)	0.910
Self-control questionnaire with answers (q.2.I.6)	0.566
Self-control MOODLE tests (q.2.I.7)	0.871
Additional exercises without answers (q.2.I.8)	0.948
Program codes of considered exercises (q.2.I.9)	0.326

In addition, to assess which form of course material students prefer and which form they consider ineffective, the boxplot was created (Figure 2). Here, the analysis was performed among all students, without taking into account their personality dimensions. The study results showed that students prioritise outlines or slides, self-control questionnaires with answers and self-control MOODLE tests (with a median of 6). The lowest scores (with a median of 1) were given to scientific literature and additional literature, such as textbooks (not scientific articles). The Wilcoxon Signed Ranks Test indicated that the choice of mentioned forms of course material (q.2.I.1, q.2.I.4, q.2.I.5, q.2.I.6, q.2.I.7) statistically significantly differ from other forms of course material as $p < 0.05$

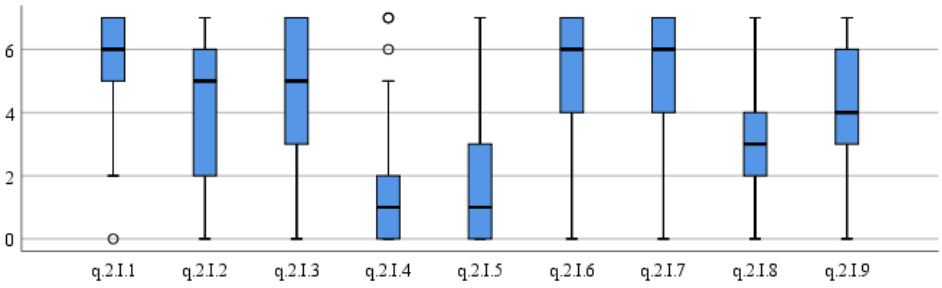


Figure 2. Boxplot diagrams of students' preferences for different forms of course materials presented in the Moodle environment (see Table 2)

Furthermore, Part II (Questions II and III) of the Questionnaire was also aimed at revealing student preferences for the length of lecture videos and the size of lecture materials (slides, textbooks). The study results showed that the majority (67.6%) of students prefer short lecture videos over a long recording of the entire lecture. The Chi-Square Test result ($p = 0.519$, $df = 2$) indicate that the preference for the length of video recordings does not depend on the student personality dimension. Also, students prioritised comprehensively presented lecture material over a concise summary of the lecture (54.5%). This preference does not depend on the student personality dimension, as confirmed by the Chi-Square Test ($p = 0.452$, $df=2$).

In Part III, Question I (see Part III, Question I, in WEB, a, Questionnaire), students had to rate the usefulness of the organisational structure of the lecture, taking into account their own level of information absorption. The p-values of all Kruskal-Wallis tests, except for “Solving tasks in student groups with the possibility to consult with the lecturer (q.3.I.5)”, exceeded 0.05 (see Table 3). Thus, students with different personality dimensions rate organisational structures in the lecture similarly, except q.3.I.5.

Table 3. Statistical significance difference between personality dimensions' groups, considering student preferences for different organisational structures of the lecture

Rate the usefulness of organising lecture structure, taking into account your own level of information absorption (from 1 – completely useless, to 7 - very useful):	**p, df =2**
Independent solving of the self-control questionnaire with answers (q.3.I.1)	0.620
Lecturer provides detailed solutions to the tasks. Student follows the lecture without active participation (q.3.I.2)	0.654
Lecturer presents material in detail, the student is constantly involved in the discussion, on issues related to the course (q.3.I.3)	0.994
Lecturer presents problematic questions and, together with the students, searches for answers to the questions of the topic being analysed (q.3.I.4)	0.760
Solving tasks in student groups with the possibility to consult lecturer (q.3.I.5)	0.046
Students solve tasks at the blackboard (q.3.I.6)	0.350
Student presents a new topic proposed by the teacher, which he analysed independently at home (q.3.I.7)	0.063

To evaluate which organisational structures of the lecture students prefer (without taking into account their personality dimensions), the boxplot diagram was assessed (see Figure 3). The results of the study showed that students rated "Independent solving of self-control questionnaire with answers (q.3.I.1)" with higher scores (with a median of 6). The lowest scores (with a median of 2) were given for "Student presents a new topic proposed by the teacher, which he analysed independent at home (q.3.I.7)".
The Wilcoxon Signed Ranks Test indicated that the mentioned organisational structures of the lecture (q.3.I.1, q.3.I.7) statistically significantly differed from other structures examined in this study ($p < 0.05$).

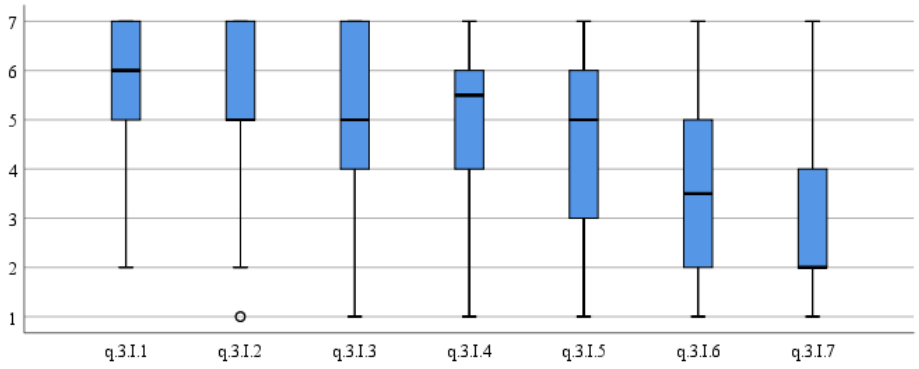


Figure 3. Boxplot diagrams of student preferences for different organisational structures of the lecture (see Table 3)

Let's recall that the p-value of the Kruskal-Wallis test for "Solving tasks in student groups with the possibility to consult with the lecturer (q.3.I.5)" is 0.046 < 0.05. For q.3.I.5, Mann-Whitney Test results showed that only choices of agreeable and conscientious students statistically significantly differ, as p = 0.023 < 0.05. Furthermore, according to the boxplot (see Figure 4), agreeable students are more likely to choose to perform tasks in student groups during the lecture with the opportunity to consult with the lecturer than conscientious students.

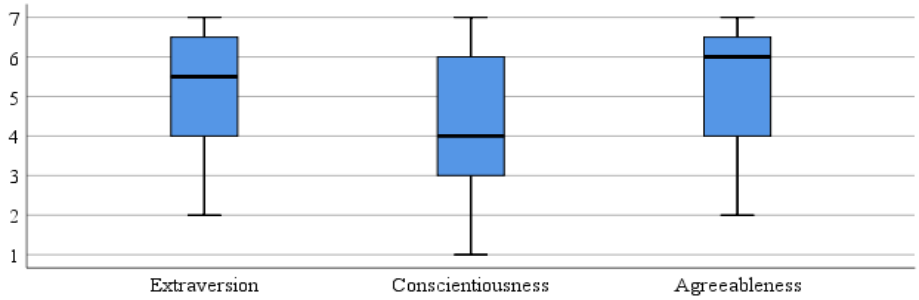


Figure 4. Boxplot diagrams of student preferences for q.3.I.5 (see Table 3)

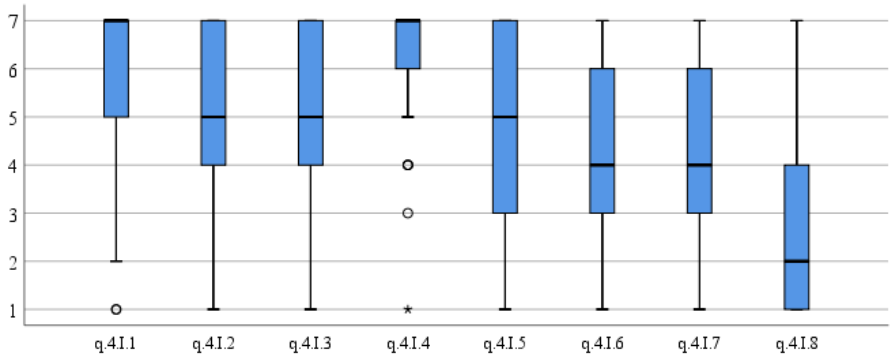
Additionally, in this part of the study (see Part III, Question II, in WEB, a, Questionnaire), students were also asked which form of lecture, in their opinion, is most helpful in mastering the topic under consideration. Students could choose one of four options: (1) when the lecturer relies on his/her own personal experience; (2) when he/she relies only on textbook material and scientific literature facts; (3) when he/she combines textbook content with examples from personal experience; (4) when the form of presentation is not significant for the student. Student answers to additional questions allowed us to determine that for the majority (91.9%) of them, it is most beneficial when the lecturer combines textbook material with examples from their personal experience.

To analyse the responses to the questions about how students evaluate the impact of the organizational study process on their own learning process (see Part IV, Question I in WEB, a, Questionnaire), it was found that the evaluations did not statistically significantly differ (student preferences are similar) considering personality dimensions, as indicated by the results of the Kruskal-Wallis test (see Table 4), since for all cases p > 0.05.

Table 4. Statistical significance difference between personality dimensions' groups, considering the influence of the organisational study process

Rate the influence of the organisational study process, taking into account your own learning process (from 1 - not at all important to 7 - very important):	**p, df =2**
At the beginning of the semester to know exact assessments days (q.4.I.1)	0.496
Review of the study course card (q.4.I.2)	0.933
To get the list of all topics of the study course in advance (q.4.I.3)	0.069
Know the assessment course in advance (q.4.I.4)	0.072
Know in advance what kind of literature will be needed for studying (q.4.I.5)	0.265
Strictly follow dates of assessments (q.4.I.6)	0.275
Strictly follow content of the course (q.4.I.7)	0.554
Strict marking of class attendance (q.4.I.8)	0.686

Furthermore, to evaluate which item of the organisational study process students prefer and which they believe negatively affects their learning process, medians were assessed (see Figure 5). Here, analysis was performed without taking into account their personality dimensions. The study results showed that students rated the opportunity to know exactly all assessment dates at the beginning of the semester (q.4.I.1) and the assessment system in advance with higher scores (q.4.I.4) (with a median of 7). The lowest score was given to the strict marking of class attendance (q.4.I.8) (with a median of 2). After conducting the Wilcoxon Signed Ranks test, it was determined that q.4.I.1, q.4.I.4, q.4.I.8 differ statistically significantly from other items examined in this study (as $p < 0.05$).



F**igure 5.** Boxplot diagrams of student preferences for different items of organizational study process (see Table 4)

The next question that was analysed is intended to understand how students evaluate the impact of the lecturer's communication style on their own learning quality (see Part V, Question I in WEB, a, Questionnaire). It was found that p-values of all Kruskal-Wallis tests, except "Lecturer who encourages reflection on students' personal progress (q.5.I.12)", exceeded 0.05 (see Table 5). Accordingly, students with different personality dimensions rated the lecturer's communication styles similarly, except for q.5.I.12.

Table 5. Statistical significance difference between personality dimensions' groups, considering student preferences for the lecturer's communication style

Rate how the lecturer's communication style affects the quality of your education (1 – strong negative influence, 2 – medium negative influence, 3 – small negative influence, 4 – no influence, 5 – small positive influence, 6 – medium positive influence, 7 – strong positive influence)	**p, df =2**
Lecturer is charismatic (q.5.I.1)	0.487
Lecturer is agreeable (q.5.I.2)	0.147
Lecturer strictly follows the rules and demands from the students (q.5.I.3)	0.723
Lecturer is restrained (q.5.I.4)	0.736
Lecturer emphasizes his status (q.5.I.5)	0.405
Lecturer is flexible (q.5.I.6)	0.939
Lecturer is supportive and willing to help (q.5.I.7)	0.551
Lecturer does not tolerate academic (q.5.I.8)	0.235
Lecturer is forgiving and gives everyone a personal opportunity, at least to receive positive evaluations (q.5.I.9)	0.455
Lecturer supports the students' initiative (q.5.I.10)	0.624
Lecturer encourages students to take responsibility (q.5.I.11)	0.774
Lecturer who encourages reflection on students' personal progress (q.5.I.12)	0.035

Furthermore, the Mann-Whitney Test results showed that the evaluations of students belonging to the extraversion personality dimension statistically significantly differ from those of other personality dimensions when assessing q.5.I.12 ($p = 0.035 < 0.05$). In addition, Mann-Whitney Test results showed that only choices of extravert and agreeable, also extravert and conscientious students significantly differed, as in both cases, $p = 0.016 < 0.05$. According to the boxplot, students who have extraversion personality dimension were more inclined than those of other personality dimensions to choose the lecturer's encouragement for students to reflect on their personal progress (see Figure 6).

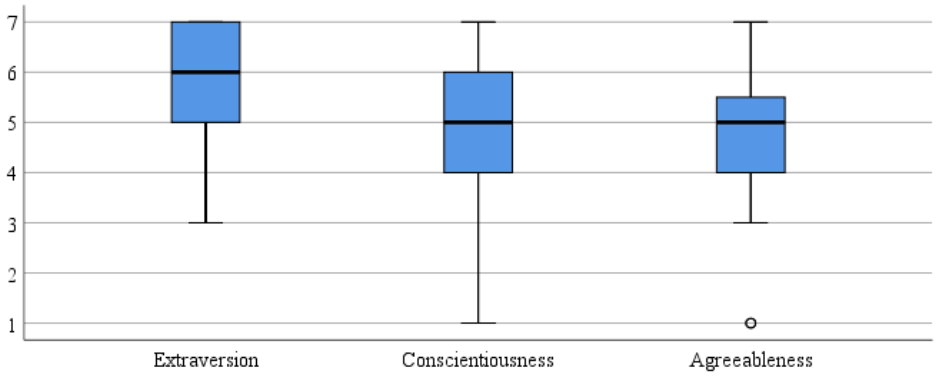


Figure 6. Boxplot diagrams of student preferences for q.5.I.12 (see Table 5)

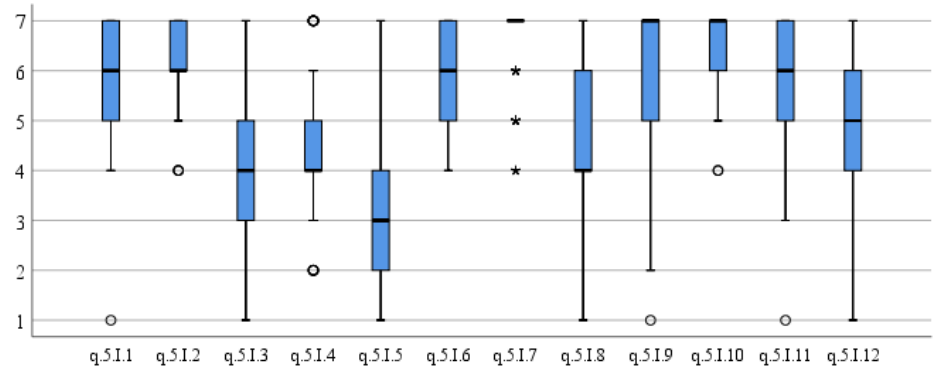


Figure 7. Boxplot diagrams of students' preferences for different lecturer's communication style (see Table 5)

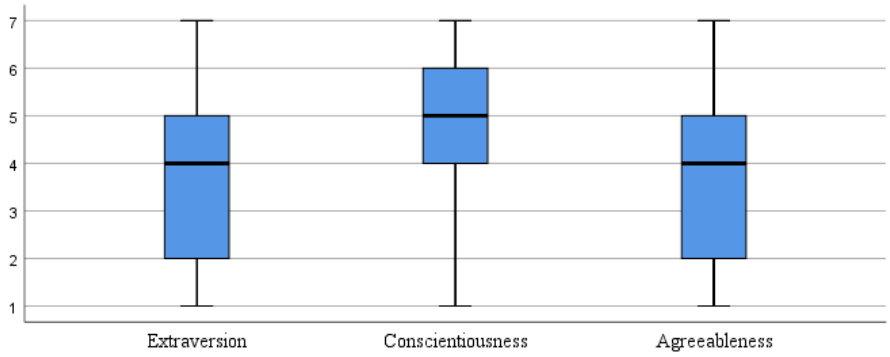
As in other parts, the medians of students' preferences were considered without taking into account their personality dimensions. The results of Boxplot diagrams (see Figure 7) showed that students rated "Lecturer is forgiving and gives everyone a personal opportunity, at least to receive positive evaluations (q.5.I.9)" and "Lecturer supports the students' initiative (q.5.I.10)" with the highest scores (with a median of 7). The lowest scores (with a median of 2) were given for "Lecturer emphasises his status (q.5.I.5)". The Wilcoxon Signed Ranks Test indicated that q.5.I.5 statistically significantly differ from other items with the lowest scores, as $p = 0.000 < 0.05$. Also, q.5.I.9 statistically significantly differ from almost all items examined in this study (as $p < 0.05$), except for q.5.I.1, q.5.I.2, q.5.I.6, as, accordingly, $p = 0.783$, $p = 0.131$, $p = 0.625$. In addition, q.5.I.10 statistically significantly differs from almost all items examined in this study (as $p < 0.05$), except for q.5.I.2, for which $p = 0.239$.

In order to examine the factors that may influence student personal learning process, 14 questions were created (see Part VI, Question I in WEB, a, Questionnaire). After conducting the Kruskal-Wallis test (see Table 6), it was found that only two questions had different student responses depending on personality dimension. Students' opinions varied regarding "To be one of the best students in the study process (q.6.I.5)" ($p = 0.029 < 0.05$) and "The opportunity to study together with someone (q.6.I.11)" ($p = 0.002 < 0.05$).

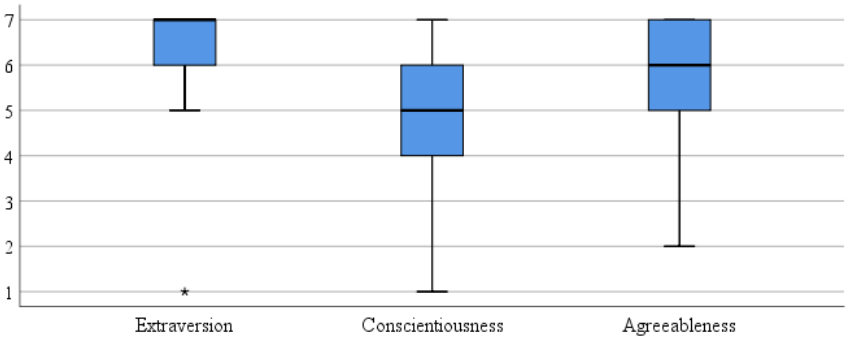
Table 6. Statistical significance difference between personality dimensions' groups, considering student preferences for different factors that influence their own learning process

Rate how important the listed factors are for your personal learning process (from 1 – not at all important, to 7 – very important):	**p, df =2**
To gain good study knowledge (q.6.I.1)	0.169
My study results would not disappoint my relatives (q.6.I.2)	0.571
My study results would not disappoint the lecturers (q.6.I.3)	0.789
Do not let yourself down in the learning process (q.6.I.4)	0.638
To be one of the best students in the learning process (q.6.I.5)	0.029
To have enough ability to study (q.6.I.6)	0.242
Be confident, not afraid of difficulties (q.6.I.7)	0.417
Reflect on my learning, what helps me and hinders my learning (q.6.I.8)	0.895
Ability to set your own learning goals (q.6.I.9)	0.744
Study based on the study goals set by the lecturer (q.6.I.10)	0.303
The opportunity to study together with someone (q.6.I.11)	0.002
The opportunity to express my opinion to the group I study with (q.6.I.12)	0.196
When it's hard to study, I try to remember what I want to achieve (q.6.I.13)	0.434
Perception of how gained knowledge can be applied in professional activities (q.6.I.14)	0.665

Moreover, analysing the responses to the questions "How important is it for you to be one of the best students in the learning process?" and "How important is it for you to have the opportunity to learn with someone?" it was found that the responses of students belonging to the dimensions of extraversion and agreeableness, as well as students belonging to personality dimensions of extraversion and conscientiousness, statistically significantly differ. The statistical significance is confirmed by the results of the Mann-Whitney Test, indicating that students belonging to the extraversion personality dimension are significantly more inclined to learn with someone ($p < 0.05$). For students belonging to conscientiousness or agreeableness personality dimensions, it is more important to be one of the best students in the learning process ($p < 0.05$) (see Figure 8).



(a) Boxplot diagrams of student preferences for q.6.I.5



(b) Boxplot diagrams of student preferences for q.6.I.11

Figure 8. Boxplot diagrams of student preferences for q.6.I.5 and q.6.I.11 (see Table 6)

In addition, from the Boxplot diagrams (see Figure 9), follows that students rated q.6.I.1, q.6.I.4, q.6.I.6, q.6.I.7, q.6.I.14 with the highest scores (with a median of 7). The lowest scores (with a median of 4) were given for q.6.I.2, q.6.I.3, and q.6.I.5. The Wilcoxon Signed Ranks Test indicated that q.6.I.1, q.6.I.4, q.6.I.6, q.6.I.7, q.6.I.14 statistically significantly differ from other examined in this study ($p < 0.05$). Mentioned items are not statistically significantly different from each other. Besides, q.6.I.3 statistically significantly differ from all other items examined in this study. q.6.I.2 is not statistically significantly different only from q.6.I.5, q.6.I.10, q.6.I.12, and q.6.I.5 is not statistically significantly different only from q.6.I.2, q.6.I.12.

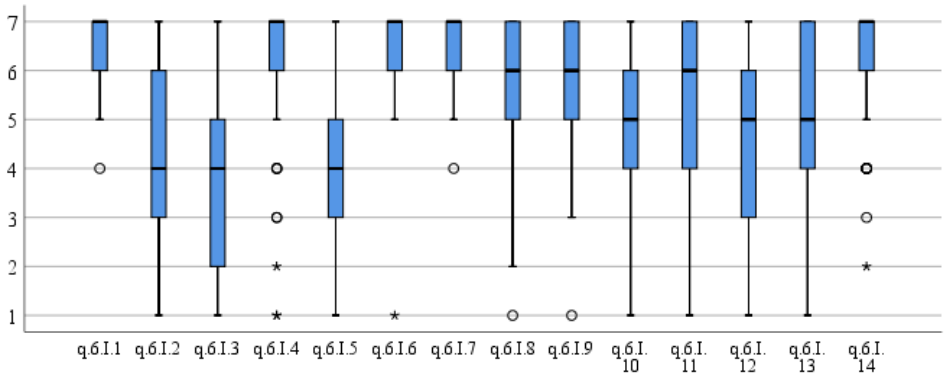


Figure 9. Boxplot diagrams of student preferences for different factors that influence their own learning process (see Table 6)

Concerning analysis for students' ability to plan time and organise their work, 10 questions were asked (see Part VII, Question I in WEB, a, Questionnaire). After analysing the response data using the Kruskal–Wallis test, it was found that, considering personality dimensions of students, the responses did not statistically significantly differ ($p > 0.05$) (Table 7). Thus, responses of students with different personality dimensions are similar.

Table 7. Statistical significance difference between personality dimensions' groups, considering student preferences for their abilities to plan time and organize their work

Rate how often it occurs during own learning process (0 - never occurs, from 1 - very rarely, to 7 - very often)	**p, df =2**
When I have a lot of work to do, I think about which work is most important and try to focus on it (q.7.I.1)	0.913
I never find time to study (q.7.I.2)	0.559
If I don't understand something, I skip it (q.7.I.3)	0.593
As soon as I sit down to study something, all kinds of distractions appear (q.7.I.4)	0.152
I create a time schedule to remember, organise and plan activities (q.7.I.5)	0.476
I stick to the schedule of tasks I have made (q.7.I.6)	0.219
I waste a lot of time trying to find information I have somewhere (q.7.I.7)	0.111
When I see that someone is not doing well in their studies, I offer to help (q.7.I.8)	0.302
When I encounter problems while studying, I usually ask for help from my fellow students or the teacher (q.7.I.9)	0.423
I know how to choose the material I need from a large number of information sources (q.7.I.10)	0.720

Additionally, from the Boxplot diagrams, it can be observed that students tend to pay more attention to planning tasks according to their importance and priorities. This conclusion can be drawn by observing medians (Figure 10) of the responses to question q.7.I.1. The lowest scores students give to the question "I never find time to study (q.7.I.2) (with a median of 3). The Wilcoxon Signed Ranks Test indicated that q.7.I.1 statistically significantly differs from other items examined in this study ($p = 0.000 < 0.05$). Also, q.7.I.2 statistically significantly differs from almost all other items examined in this study, except for q.7.I.3, q.7.I.5, q.7.I.6, as accordingly $p = 0.259$, $p = 0.108$, $p = 0.120$.

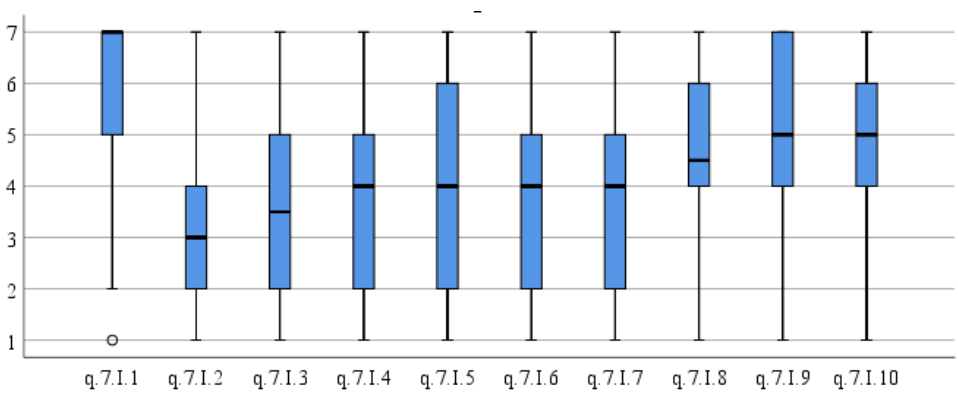


Figure 10. Boxplot diagrams of student preferences for their abilities to plan time and organise their work (see Table 7)

4. Conclusions

A study conducted with informatics engineering students first showed that the organisational structure, the forms of materials, the lecture-related tools, and the lecturer's communication style affect these students' learning. These students are representatives of Gen Z, for whom the quality of studies and an understanding of how the acquired knowledge is applied in professional activities are important.

According to the analysis, the majority of respondents (97.83%) studied in a state-funded place, and most of them (70.65%) studied in their dream program. Thus, the majority of students at the beginning of studies are motivated to learn, and the lecturer becomes responsible for organising a learning strategy, in order to ensure the needs of most students.

There is an ongoing debate regarding the most appropriate psychological framework for categorising personality. However, the most common model in the literature is the Big Five model. In this model, each personality is considered to be associated with five main dimensions: neuroticism, extraversion, openness to experience, agreeableness and conscientiousness.

In the current study only the personality dimensions of agreeableness, conscientiousness, and extraversion possessed a sufficient sample size for further data

analysis. Study results showed that for conscientious or agreeable students, being among the top students in the academic process is significantly essential. Also, analyses of scientific literature reveal that conscientiousness and openness were the most significant predictors of academic performance. Corresponding to the considered research papers, the connection between learning success and personality dimensions may vary slightly across different academic fields. Thus, this study is planned to be repeated in other study programs where mathematics courses are taught.

Regarding the study results, there is no statistically significant difference between personality dimensions and student preferences for most learning strategies. However, factors related to their working within student groups, becoming one of the best students in the learning process, and their expectation of encouragement for reflection on their own personal progress had a statistically significant difference with their personality dimensions.

The study's results indicated that students prefer to learn using short videos and lecture slides rather than course textbooks, scientific articles, or long lecture videos. Besides, students prioritise outlines or slides, self-control questionnaires with answers and self-control MOODLE tests. For the majority (91.9%) of students, it is most beneficial when the lecturer combines textbook material with examples from their personal experience.

Furthermore, the study results showed that it is essential for students to be fully aware of all due dates and the assessment system at the start of the semester. For students who exhibit the extraversion personality dimension, receiving feedback from a teacher regarding their academic progress is more crucial than for others. These students tend to study with friends or in groups. Also, as mentioned earlier, for conscientious or agreeable students, being among the top students in the academic process is significantly important. Additionally, it was observed that students tend to pay more attention to planning tasks according to their importance and priorities.

The study is planned to be replicated, involving more student groups from different study programs, since in the current study, only the personality dimensions of agreeableness, conscientiousness, and extraversion possessed a sufficient sample size for further data analysis. Also, the study is planned to be conducted among different faculties that have the same mathematics courses. Taking into account the offer of new learning methods, it is planned to conduct the study by including newly applicable methods in the questionnaire.

References

Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *Internet and Higher Education*, **45**, 100724. https://doi.org/10.1016/j.iheduc.2019.100724

Afshar, M. A., Jafari, A., Heshmati, F., Movahedzadeh, F., Cherif, A. H. (2019). Instructional strategies for motivating and engraining generation z students in their own learning process. *Journal of Education and Practice*, **10(3)**, 1-19.

Anglim, J., & O'Connor, P. (2019). Measurement and research using the Big Five, HEXACO, and narrow traits: A primer for researchers and practitioners. *Australian Journal of Psychology*, **71(1)**, 16–25. https://doi.org/10.1111/ajpy.12202

Anthony, G. (2000). Factors influencing first-year students' success in mathematics. *International Journal of Mathematical Education in Science and Technology*, **31(1)**, 3–14. https://doi.org/10.1080/002073900287336

Baruth, O., Cohen, Anat. (2023). Personality and satisfaction with online courses: The relation between the Big Five personality traits and satisfaction with online learning activities. *Education and Information Technologies*, **28**, 879–904. https://doi.org/10.1007/s10639-022-11199-x

Biwer, F., Egbrink, M. G. A. o., Aalten, P., de Bruin, A. B. H. (2020). Fostering effective learning strategies in higher education—A mixed-methods study. *Journal of Applied Research in Memory and Cognition*, **9(2)**, 186–203. https://doi.org/10.1016/j.jarmac.2020.03.004

Bligh, D., Cameron, B. J. (2000). What's the use of lectures?. *The Canadian Journal of Higher Education*, **30(1)**, 192.

Bunevicius A. (2005). Didžiojo penketo asmenybės dimensijos (DPAD). *Biologinė Psichiatrija ir Psichofarmakologija*, **6**, p. 34.

Cakula, S. (2024). Digital learning model. *Baltic J. Modern Computing*, **12 (3)**, 248–258 https://doi.org/10.22364/bjmc.2024.12.3.02

Çağataylı, M., Çelebi, E. (2022). Estimating Academic Success in Higher Education Using Big Five Personality Traits, a Machine Learning Approach. *Arabian Journal for Science and Engineering*, **47(2)**, 1289–1298. https://doi.org/10.1007/s13369-021-05873-4

Cilliers, E. J. (2022). The challenge of teaching generation Z. *PEOPLE: International Journal of Social Sciences*, **3(1)**, 188–198. https://doi.org/10.20319/PIJSS.2017.31.188198

Dolot, A. (2018). The characteristics of Generation Z. *E-Mentor*, **74(74)**, 44–50. https://doi.org/10.15219/em74.1351

Durak, Y. H. (2023). Role of personality traits in collaborative group works at flipped classrooms. *Current Psychology*, **42(15)**, 13093–13113. https://doi.org/10.1007/s12144-022-02702-1

Feher, A., Vernon, P. A. (2021). Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality. *Personality and Individual Differences,* **169**, Article 110002. https://doi.org/10.1016/j.paid.2020.110002

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology*: Vol. 2 (pp. 141-165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, **59(6)**, 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldberg L. R. (1999) A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*. **7**, 7–28.

Hendrastomo, G., Januarti, N. (2023). The Characteristics of Generation Z Students and Implications for Future Learning Methods. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran*, **9(2)**, 484-496. https://doi.org/10.33394/jk.v9i2.7745

Hernandez-De-Menendez, M., Escobar Díaz, C. A., Morales-Menendez, R. (2008). *Educational experiences with Generation Z*. **14**, 847–859. https://doi.org/10.1007/s12008-020-00674-9

Jensen, M. (2015). Personality Traits, Learning and Academic Achievements. *Journal of Education and Learning*, **4(4)**, 91. https://doi.org/10.5539/jel.v4n4p91

Juste, M. P., López, B. R. (2010). Learning strategies in higher education. International journal of Learning, **17(1)**, 259-274. https://doi.org/10.18848/1447-9494/CGP/v17i01/46813

Kamal, A., Radhakrishnan, S. (2019). Individual learning preferences based on personality traits in an E-learning scenario. *Education and Information Technologies*, **24(1)**, 407–435. https://doi.org/10.1007/s10639-018-9777-4

Komarraju, M., Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and Individual Differences*, **39(3)**, 557–567. https://doi.org/10.1016/j.paid.2005.02.013

Koner, S., Mazumder, S. (2024). Academic motivation, self-esteem and personality factors: A review of effect on mathematics anxiety. *International Journal of Literacy and Education*, **4(1)**, 77-86. https://doi.org/10.22271/27891607.2024.v4.i1a.167

Kumari, D. R., Tiwari, D. R., Akhtar, R., Gupta, S. (2023). Traditional teaching method Vs Modern teaching method. *GPH-International Journal of Educational Research*, **6(11)**, 80-85. https://doi.org/10.5281/zenodo.10429275

Lopez, E. N., Abadiano, M. N. (2023). Understanding Generation Z, the new generation of learners: a technological-motivational-learning theory. *Journal of Harbin Engineering University*, **44(10)**, 770-784.

Lyons, M., Limniou, M., Schermbrucker, I., Hands, C., Downes, J. J. (2017). The big five, learning goals, exam preparedness, and preference for flipped classroom teaching: Evidence from a large psychology undergraduate cohort. *Psychology Learning and Teaching*, **16(1)**, 36–46. https://doi.org/10.1177/1475725716680460

Marin, K. A., White, S. J. (2023). Generation Z goes to math class: How the effective mathematics teaching practices can support a new generation of learners. *School Science and Mathematics*, **123(1)**, 31–37. https://doi.org/10.1111/ssm.12565

McCrae, R. R., Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, **52(1)**, 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCredie, M. N., Kurtz, J. E. (2020). Prospective prediction of academic performance in college using self- and informant-rated personality traits. *Journal of Research in Personality*, **85**, 103911. https://doi.org/10.1016/j.jrp.2019.103911

Muhid, A., Ridho, A., Yusuf, A., Wahyudi, N., Ulya, Z., Asyhar, A. (2021). Big five personality test for state islamic senior high school students in indonesia. *International Journal of Instruction*, **14(2)**, 783-500. https://doi.org/10.29333/iji.2021.14227a

Munna, A. S., Kalam, M. A. (2021). Teaching and learning process to enhance teaching effectiveness: a literature review. *International Journal of Humanities and Innovation (IJHI)*, **4(1)**, 1–4. https://doi.org/10.33750/IJHI.V4I1.102

Opoku, O. G., Adamu, A., Daniel, O. (2023). Relation between students' personality traits and their preferred teaching methods: Students at the university of Ghana and the Huzhou Normal University. *Heliyon*, **9(1)**, e13011. https://doi.org/10.1016/j.heliyon.2023.e13011

Pliuskuvienė, B., Radvilaitė, U., Juodagalvytė, R., Ramanauskaitė, S., Stefanovič, P. (2024). Educational data mining and learning analytics: Text generators usage effect on students' grades. *New Trends in Computer Sciences*, **2(1)**, 19-30. https://doi.org/10.3846/ntcs.2024.21318

Rajapakshe, W. (2017). *A Study on the Big Five Personality Dimensions' Effect on University Students' Academic Performance*. **19**, 69–75. https://doi.org/10.9790/487X-1912016975

Smaliukiene, R., Kocai, E., Tamuleviciute, A. (2020). Generation Z and consumption: How communication environment shapes youth choices. *Medijske Studije*, **11(22)**, 24–45. https://doi.org/10.20901/ms.11.22.2

Sugano, S. G. C., Mamolo, L. A. (2021). The Effects of Teaching Methodologies on Students' Attitude and Motivation: A Meta-Analysis. International Journal of Instruction, **14(3)**, 827-846. https:// doi.org.10.29333/iji.2021.14348a

Twenge, J. M. (2017). *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy--and completely unprepared for adulthood--and what that means for the rest of us*. Simon & Schuster.

Wakefield, J., Tyler, J. (2023). Does active participation via integrated questions in large lectures matter? *Heliyon*, **9(5)**, e15873. https://doi.org/10.1016/j.heliyon.2023.e15873

Vedel, A., Poropat, A. E. (2017). Personality and Academic Performance. In *Encyclopedia of Personality and Individual Differences* (pp. 1–9). https://doi.org/10.1007/978-3-319-28099-8_989-1

Voskoglou, M. (2019). Comparing teaching methods of mathematics at university level. *Education Sciences*, **9(3)**. https://doi.org/10.3390/educsci9030204

WEB (a). Questionnaire available at https://forms.gle/v1NYkonTmYTajQXX6

Received March 11, 2025, revised February 1, 2026, accepted February 28, 2026

Baltic J. Modern Computing, Vol. 14 (2026), No. 1, 238-261
https://doi.org/10.22364/bjmc.2026.14.1.13

Preliminary Reassessment Results of the Recent Vertical Movements of the Earth's Crust in Bulgaria

Vasil CVETKOV, Christina MICKRENSKA

University of Architecture, Civil Engineering and Geodesy,
Hristo Smirnenski Blvd. 1, Sofia, Bulgaria

tzvetkov_vasil@abv.bg, kmikrenska_fgs@uacg.bg

ORCID 0000-0001-9628-6768, ORCID:0000-0002-7054-1799

Abstract. Recent vertical crustal movements in Bulgaria have been mapped using adjusted levelling data from three epochs, yet results from different research teams are inconsistent and often contradict known tectonic fault structures. To provide an independent evaluation, we applied a new estimation approach based on data from Bulgaria's Second (1953–1957) and Third (1975–1980) levelling campaigns. Both networks were processed through 3^n independent adjustments, selecting line elevations that minimised loop misclosures. In addition, Inverse Distance Weighting (IDW) with a power parameter of 6 was applied. This procedure reduced benchmark height standard errors to below ±5.3 mm and vertical velocity errors to ±0.30 mm/year. The resulting high-accuracy map reveals a strong correlation between tectonic boundaries and major earthquake epicentres.

Keywords: Probability Theory, Recent Movements of Earth's Crust, Geometric Levelling, Adjustment Algorithms, Earthquake Hazards

1. Introduction

1.1. Tectonic scheme of Bulgaria

Bulgaria lies within two major tectonic zones: the northern part of the Alpine thrust belt in the Balkans and its foreland, the Moesian Platform. Bulgarian geologists widely recognise these units, but debates persist over how to divide them further, interpret their structures, and understand their geodynamic history.

Following the widespread acceptance of plate tectonic theory around forty years ago, a variety of competing models have been proposed to interpret Bulgaria's tectonic structure and geological evolution. Most of them are built on the idea that the Alpine development of the Balkans was shaped by ongoing subduction and multiple continental collisions along Eurasia's southern margin. While these models differ in how they define tectonic units, they generally agree on one point: the Alpine thrust belt is a patchwork of both local and exotic continental fragments that accreted to Eurasia during the closure of the Tethys Ocean.

Despite this long-standing interest, no up-to-date tectonic map of Bulgaria has been published. Developing a reliable tectonic model requires a comprehensive analysis of geological data, including lithology, stratigraphy, structure, paleontology, volcanism, plutonism, metamorphism, geochronology, and geophysics. This is the approach the authors (Dabovski, C., Boyanov, I., Khrischev, K., Nikolov, T., Sapounov, I., Yanev, Y., and Zagorchev, I., 2002, Figure 1) have taken in creating a new model for the Alpine structure and evolution of Bulgaria, shown in Figure 1.

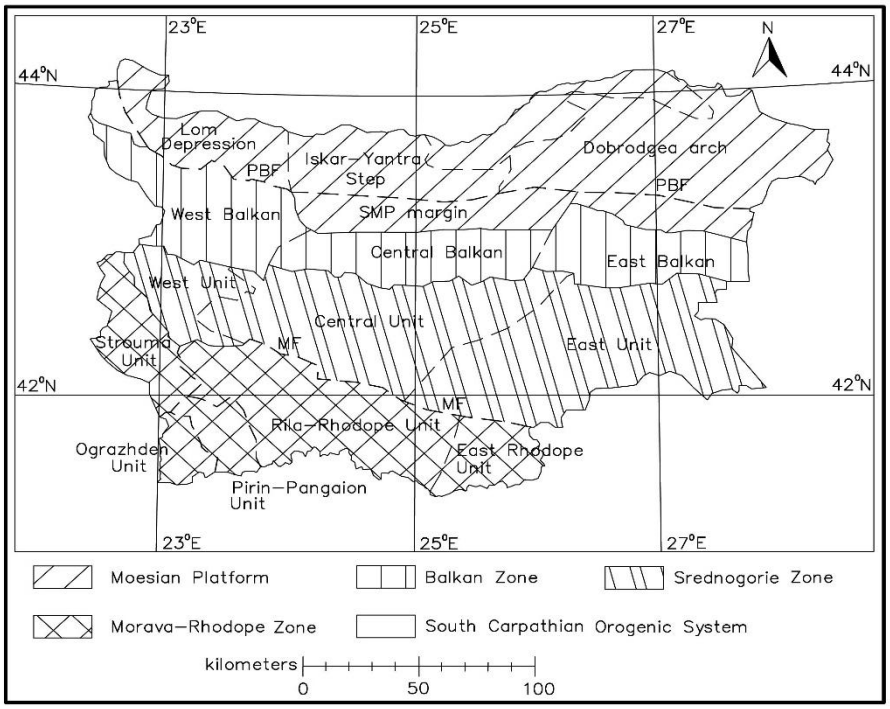


Fig. 1. Tectonic Scheme of Bulgaria. (Based on Dabovski, C., Boyanov, I., Khrischev, K., Nikolov, T., Sapounov, I., Yanev, Y. and Zagorchev, I., 2002, Figure 1)

In this model, tectonic units are understood as rock bodies that are confined dominantly between fault structures. Each unit is characterised by a specific rock assemblage (specific lithology, stratigraphy, magmatism, metamorphism) and deformational events (age, resulting structures), recording a geological evolution different from that of the adjacent units (Dabovski, C., Boyanov, I., Khrischev, K., Nikolov, T., Sapounov, I., Yanev, Y. and Zagorchev, I., 2002, Figure 1).

1.2. Kanev and Mladenovski's map from 1969

One of the most popular maps of the recent vertical movements and their speeds in the territory of Bulgaria was created by Kanev and Mladenovski (Kanev and Mladenovski, 1969). The map, shown in Figure 2, is based on the results from the First (1920-1930) and the Second (1953-1957) Levelling campaigns in our country.

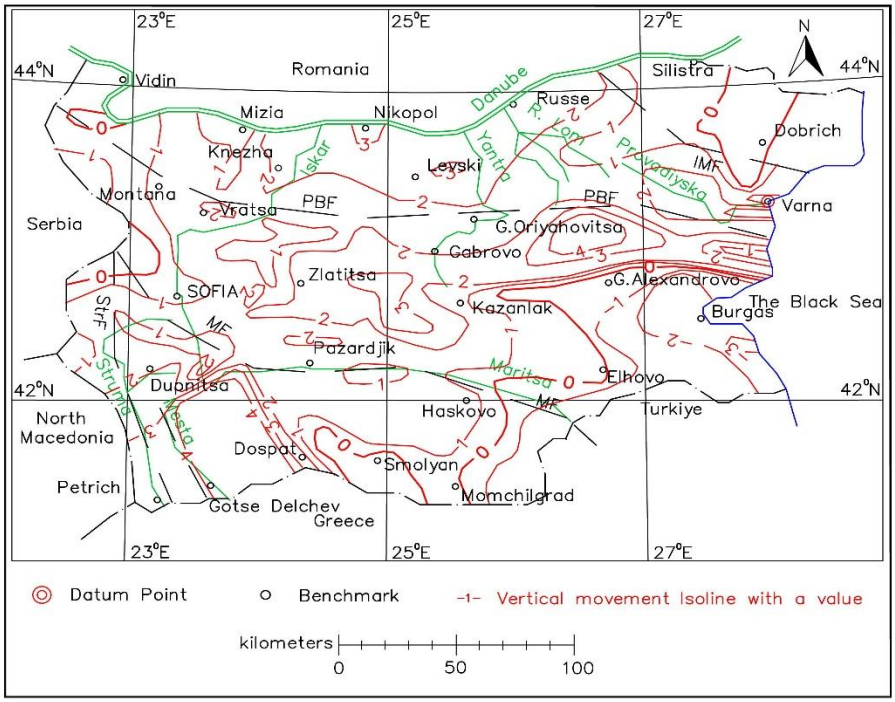


Fig. 2. Kanev and Mladenovski's map from 1969. (Kanev, D. and Mladenovski, M., 1969, Figure 2)

According to Figure 2, almost the whole territory of Bulgaria is rising. The magnitude of the uplifts is mainly 1-2 mm/y. The supreme uplift velocities, with values exceeding 4 mm/y, are in the Pirin mountain and the Kotel-Omurtag part of the Balkan /Stara Planina/. The sinking areas are the South-Middle Rhodopes and the Strandja–Sakar mountains, where the drop speeds of the Earth's crust are up to -3 mm/y. The standard errors of the estimated vertical speeds used weights and additional details regarding the adjustment of the levelling networks are not given in the publication (Kanev and Mladenovski, 1969).

1.3. Spiridonov and Georgiev's map from 2003

Another assessment of recent vertical crustal movements in Bulgaria is provided by Spiridonov and Georgiev (2003). Their results indicate subsidence within the Lom Depression at a rate of approximately −2.5 mm/year, whereas the central Moesian Plain exhibits uplift of about 1 mm/year. The Western Forebalkan and much of the Balkan region are subsiding at approximately -1 mm/year, whereas the Central Balkan is rising at 2 mm/year. Uplift rates for the Rila, Pirin, Rhodopes, and Sakar mountains are 2 mm/year, 4 mm/year, 3.5 mm/year, and 1.5 mm/year, respectively. The study does not provide standard errors for these values. The authors themselves note a lack of geomorphological consistency between their results and the pattern of known active faults in Bulgaria. Notably, the South-Moesian fault [7, 8] is absent from their findings.

1.4. Belyashki's map from 2012

A different picture of the tectonic motions in the territory of Bulgaria is presented by Belyashki (Belyashki, T., 2012, Figure 3). Using the first and the second order levelling data from the First and the Third Levelling campaign in Bulgaria, Belyashki created the map presented in Figure 3.

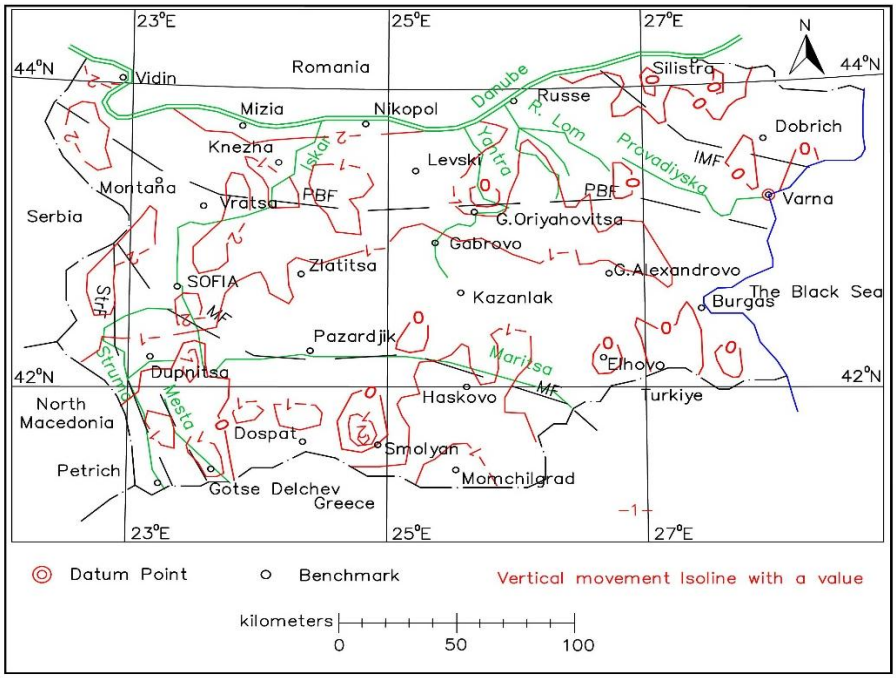


Fig. 3. Belyashki's map from 2012. (Belyashki, T., 2012, Figure 3)

The Belyashki's approach was to estimate relative speeds between line terminal benchmarks on the first stage. In the second stage, he used relative speeds to close loops, which were adjusted using weights (1).

$$w = 100.\Delta t^2/[40^2.L.(m_I^2 + m_{III}^2)] \quad (1)$$

In equation (1), L is the length of the levelling line, and m_I and m_{III} are the a posteriori accuracy of the First and the Third Levelling Campaigns.

According to Figure 3, the velocities are predominantly in the range from 0 to -2 mm/y. The velocities between -2 mm/year and -3.5 mm/year are in the Lom depression and the Western Forebalkan. The mountains Pirin, Rila, and Central Rhodopes are rising by 1-2 mm/y. A similar rise is also detected in the Lodogorie-Dobrudzha Swell and the Strandzha

mountain. The standard errors of the obtained velocities vary from 0.1 mm/year in Eastern Bulgaria to 1 mm/year in Western Bulgaria, with an average of 0.30 mm/year.

1.5. Gospodinov's team map from 2022

The most recent map of vertical crustal movement in Bulgaria was published by Gospodinov et al. (2022). That study presented three different representations of vertical tectonic motion, each varying in content, conclusions, and accuracy. The most reliable of them, based on the reported standard deviations of benchmark vertical speeds, is the map derived from levelling data collected during Bulgaria's First and Third national levelling campaigns. This version is shown in Figure 4 below.

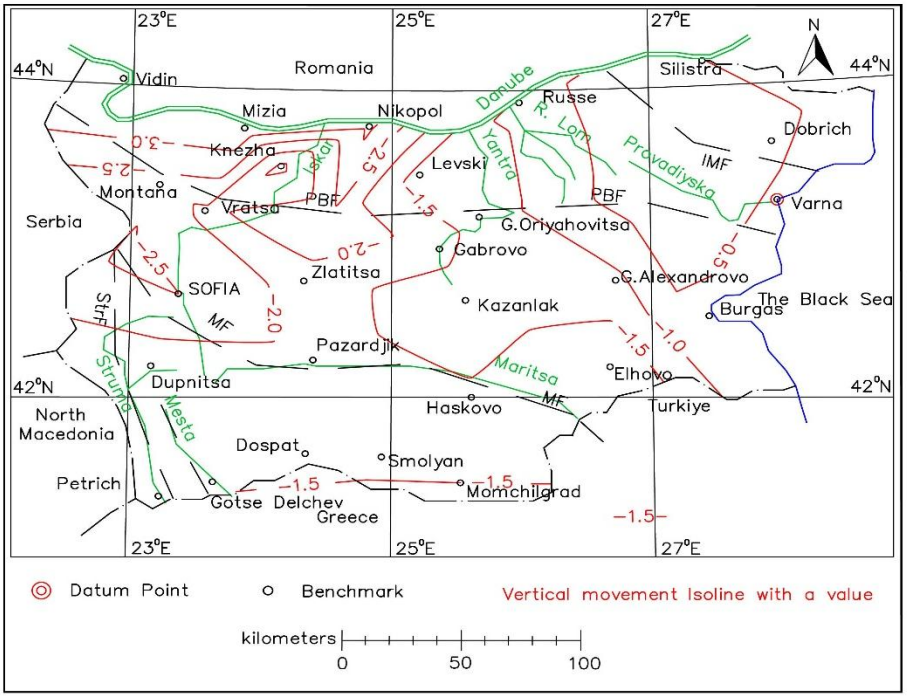


Fig. 4. Gospodinov's team map from 2022. (Gospodinov et al., 2022, Table 4, Column 7)

The applied approach is the simultaneous adjustment of networks from both epochs. The used weights are in the form (2).

$$w = 1/[\mu^2.L + \sigma^2.L^2] \quad (2)$$

In equation (2), L is the length of the levelling line, and μ and σ are the accidental and systematic errors estimated by the concrete line.

Contrary to Kanev and Mladenovski, Gospodinov's team (Gospodinov et al., 2022, Table 4, Column 7) estimates that the whole territory of Bulgaria is sinking. According to the authors, the values of the velocities are mainly in the range -1 ÷ -2 mm/y. The maximum fall of -3.4 mm/year is in the Lom depression, located in the southwest part of

the Moesian Platform. The standard errors of the determined velocities are 0.70 mm/year on average.

1.6. Research objectives

The common issues in the vertical velocity estimates from the studies above are:

- They used the mean of paired line elevation measurements during levelling adjustments - a major source of systematic error in levelling data processing (Cvetkov, 2022, 2024a, 2024b).
- Adjustments were made without minimising the standard errors of the adjusted benchmark heights (Cvetkov and Gospodinov, 2023).
- Adjustments used weights inversely proportional to the line weights. According to newly revealed facts, there is no clear and statistically proven relationship between the levelling lines and line elevation accuracy (Cvetkov, 2022, 2023, 2024c).
- Using the same data, these studies produced inconsistent outcomes that don't align with Bulgaria's known active tectonic faults (Dabovski et al., 2002; Spiridonov and Georgiev, 2003; Zagorchev, 2009; Stanciu and Ioane, 2023).

So, we define our main goal as a reassessment of the recent vertical crustal movements in Bulgaria using updated approaches grounded in recent developments in probability theory (Cvetkov, 2024a, 2024b, 2024c), modern geometric levelling adjustment methods (Cvetkov, 2024c; Cvetkov and Gospodinov, 2023), and sophisticated tests for analysing benchmark vertical displacements and their speeds (Caspary, 2000; Hamza et al., 2020; Niemeier, 1981; Savšek-Safić et al., 2006; Setan and Singh, 2001; among others).

2. Data and its processing

2.1. Used data

Appendices B and C contain the levelling data for the Second Levelling (1953–1957) and the Third Levelling of Bulgaria (1975–1984), respectively, used in this study. Figure 5 shows the configurations of the networks, along with the corresponding measurement years. The total lengths are 3,127.47 km for the Second Network and 3,568.33 km for the Third. It must be remarked that the nodal benchmarks in both networks were selected to meet the following criteria:

- Common for both networks.
- Mounted on massive public buildings such as railway stations, post offices, administrative buildings, churches, etc., or on rocks.
- All common benchmarks were estimated as stable by checking their stability against at least three neighbouring benchmarks.

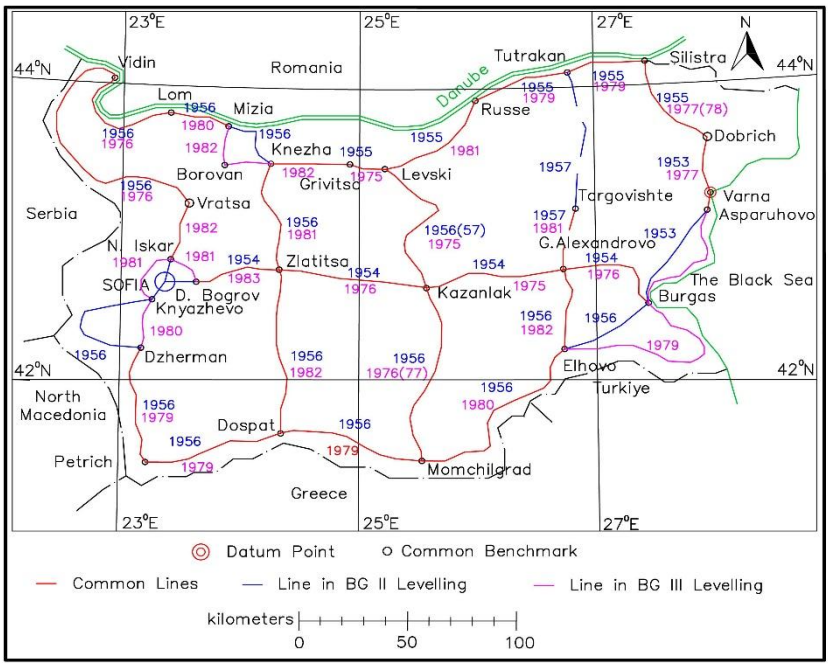


Fig. 5. Configuration of the Second Levelling Network /1953-1957/ and that part of the Third Levelling Network /1975-1984/ of Bulgaria, used in this analysis.

2.2. Data processing

Figure 5 shows that shared routes between the two levelling epochs predominate over routes unique to each epoch. This overlap improves the reliability and objectivity of the estimated vertical displacements and velocities (Caspary, 2000; Hamza et al., 2020; Niemeier, 1981; Savšek-Safić et al., 2006; Setan and Singh, 2001). Still, there are notable differences in the network configurations. The Second Levelling was primarily conducted between 1955 and 1956, while the relevant portion of the Third Levelling occurred between 1975 and 1983. During that time, some benchmarks shifted, resulting in a reduction in the final network accuracy.

The equipment and methods also varied between epochs. The Second Levelling used spirit levels - primarily Zeiss III and Ni004. The Third Levelling relied on automatic levels with compensators, using both first-order Ni002 and, after 1980, second-order Ni007 instruments. Measurement procedures also differed: the maximum sight length was 50 m in the Second Levelling, but reduced to 40 m in the Third.

Given these variations, a more advanced and refined analysis of the levelling data is justified (Caspary, 2000; Hamza et al., 2020; Niemeier, 1981; Savšek-Safić et al., 2006; Setan and Singh, 2001). The processing algorithm used is described below.

2.2.1. Selecting those observations in levelling lines that minimise the loop closing errors by a separate 3^n adjustments of each epoch network

The major drawback of the classical adjustment method for precise levelling networks is the use of the arithmetic mean of the two measured height differences between the endpoints of a levelling line. Of course, suppose we assume that both measurements of each height difference follow the same distribution with a standard deviation of σ. In that case, the arithmetic mean of the two measurements will have a standard deviation of 0.707 σ. However, this fact does not mean that the true error of the arithmetic mean of the two measurements is smaller than the true error of either individual measurement. In reality, the probability that the true error of the arithmetic mean of two random observations is smaller than the true errors of both observations converge to 33% if the distribution of the observations is uniform. In the case of normally distributed observations, this probability is even lower, under 30% (Cvetkov, 2022, 2024a, 2024b). Figure 6 illustrates this point using the closing errors in the loops from Bulgaria's Third Levelling. The procedure of 3^n adjustments is widely discussed in (Cvetkov, 2024c, 2024d). Appendix A presents the pseudocode for the 3^n selection procedure, which was implemented in a C# language program to perform the computations in this research.

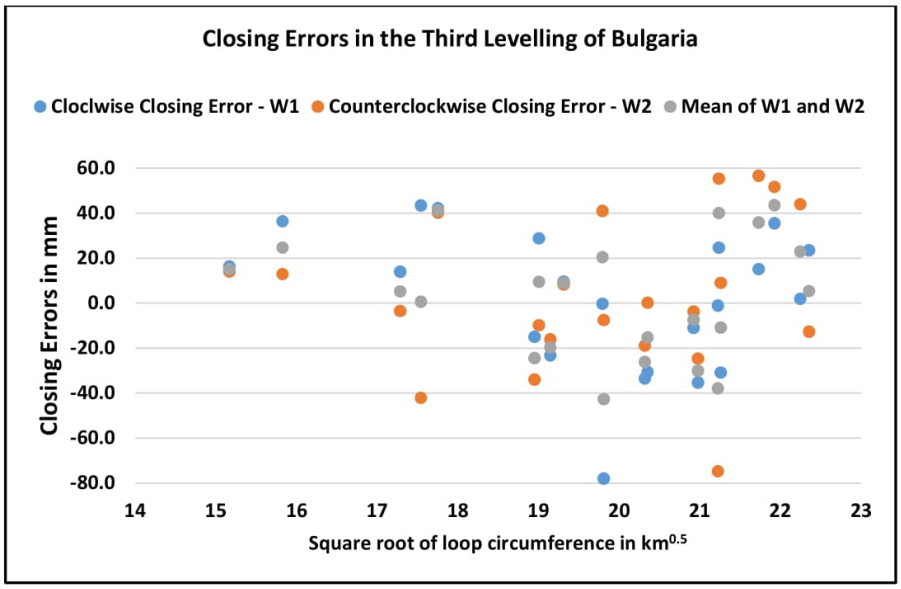


Fig. 6. Loop closing errors in the Third Levelling Network /1975-1984/ of Bulgaria.

Figure 6 shows that in over 70% of cases, the mean of observations W1 and W2 has a larger true error than either W1 or W2 individually. This suggests that using the observation with the smallest true error, rather than their mean, is a more robust and reliable method, especially in the presence of gross errors, outliers, and systematic errors, which are common in high-precision geometric levelling.

2.2.2. Minimising the outlier effects and network configuration influence by applying IDW to the networks

After selecting the most reliable observations, we carried out the final adjustments for both the Second and Third Levelling Networks. To minimize the impact of network configuration on the results, we employed iterative reweighting using IDW. The weights were defined according to equation (3).

$$w = (150(km)/L(km))^p \tag{3}$$

In equation (3), L denotes the length of levelling lines and p the power parameter. This approach further reduces the standard errors of the adjusted benchmark heights, resulting in improved accuracy of the vertical displacement estimates (Cvetkov and Gospodinov, 2023). As illustrated in Figure 7, the intersection of the summed benchmark standard errors, defined as the mean standard errors derived from the benchmark standard error samples of the Second and Third Bulgarian Levelling, occurs at a power parameter of approximately p ≈ 6. Consequently, a value of p = 6 was adopted in this study. This approach is essential not only for achieving closer alignment of the central tendencies of the benchmark standard error samples but also for ensuring greater sample homogeneity, which is critically important in displacement analysis. The homogeneity of the benchmark standard errors obtained in both measurement epochs is examined in Subsection 2.2.3.

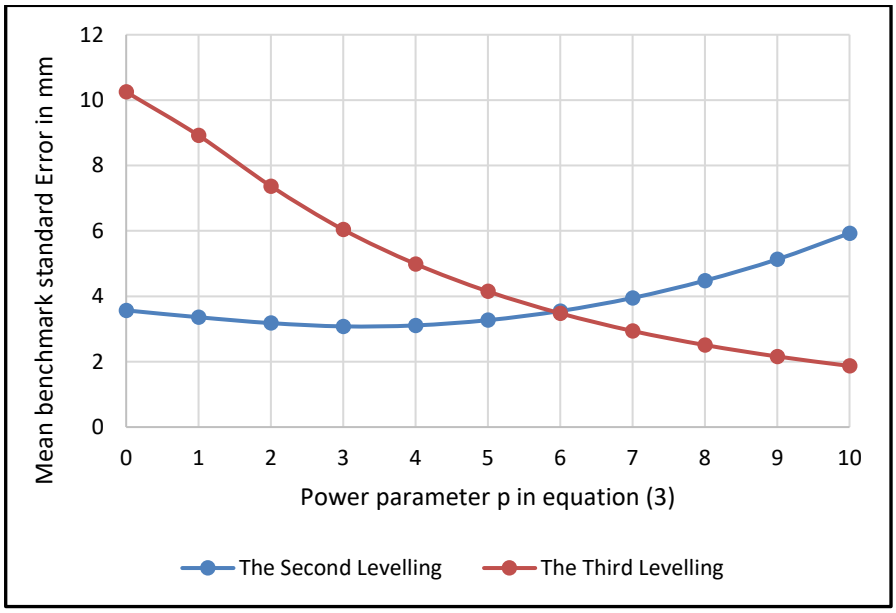


Fig.7. Mean values of benchmark's standard errors regarding the power parameter *p* in equation (3).

Appendix D presents the adjusted benchmark heights for both measurement epochs together with the derived benchmark vertical velocities (mm/year), calculated using IDW

power parameters p = {4, 5, 6, 7, 8}. The results show that selecting p within this range has no significant effect on the computed benchmark velocities. Consequently, the delineation of the principal tectonic zones remains stable under small variations of the IDW power parameter.

2.2.3. Test the a posteriori variance factors of both epochs for their compatibility

When identifying displacements, network reliability and sensitivity are critical. Therefore, it's essential to detect any hidden gross errors and ensure uniform accuracy across both networks. These conditions can be evaluated by comparing the a posteriori variance from both adjustments (Caspary, 2000; Hamza et al., 2020; Niemeier, 1981; Savšek-Safić et al., 2006; Setan and Singh, 2001). The hypotheses for this test are the null hypothesis (H_0) and the alternative hypothesis (H_a).

$$H_0: \sigma_1^2 = \sigma_2^2 \tag{4}$$

$$H_A: \sigma_1^2 \neq \sigma_2^2 \tag{5}$$

Verification of the null hypothesis can be done by the test statistic T (6):

$$T = \frac{\sigma_1^2}{\sigma_2^2} < F_{1-\alpha, f_1, f_2} \tag{6}$$

$$f_i = n_i - k_i - d_i \tag{7}$$

α–risk level,
f_i – degree of freedom of a certain epoch,
n_i – number of observations of a certain epoch,
k_i – number of unknowns of a certain epoch,
d_i – datum defect of a certain epoch.

In our final adjustments, we used 20 lines defined by the following benchmarks: Varna, Burgas, Gorno Alexandrovo, Elhovo, Tutrakan, Kazanlak, Levski, Momchilgrad, Dospat, Zlatitsa, Knezha, Vidin, Petrich, and Sofia. In the case of the Third Levelling, the nodal benchmark in Sofia was selected in Novi Iskar. The century benchmark 28 in Varna was adopted as the datum point, with an estimated vertical velocity of ±0.00 mm/year (Belyashki, 2012). According to a FIG report (Sacher et al., 2004), the height of benchmark 28 in Varna had changed by only 2 mm over 25 years (from 1958 to 1983). In addition, this benchmark was selected to maintain consistency with previous network adjustments (Gospodinov et al., 2022; Belyashki, 2012; Kanev and Mladenovski, 1969), which report its long-term stability. Thus, we have that $n_1=n_2=20$, $k_1=k_2=13$, $d_1=d_2=0$, or $f_1=f_2=7$.

If the null hypothesis (4) is not rejected, then the accuracy of both networks is compatible. There are no gross errors in observations. There is a significant similarity in the network configurations. Thus, the assessment of vertical displacements will produce unbiased and reliable estimates.

If the null hypothesis (4) is rejected, further analysis must be halted. At this point, a deeper investigation is required - reassessing the applied weights, identifying undetected

outliers, and examining differences in network configuration (Caspary, 2000; Niemeier, 1981; Savšek-Safić et al., 2006).

2.2.4. Estimation of the speed of vertical displacements and its accuracy

Before estimating the rate of vertical crustal displacements and their annual speeds, we first need to determine the central epochs of the two levelling campaigns - specifically, the period Δt between them. This is calculated as a weighted average of the measurement years, using the total length of levelling lines recorded each year as weights. For the Second Levelling Network, the weighted average year is 1955.22; for the Third, it is 1978.44. Therefore, the period is Δt = 1978.44 − 1955.22 = 23.22 years.

Suppose that $\sigma_{H_t}^2$ and $\sigma_{H_{t+\Delta t}}^2$ are the variances of a random nodal benchmark in the networks of the Second and Third Levelling, and H_t and $H_{t+\Delta t}$ are the adjusted heights of this benchmark in the epochs 1955.22 and 1978.44, the displacement of this benchmark and the accuracy of the displacement can be given by equations (8) and (9).

$$\Delta = H_{t+\Delta t} - H_t \tag{8}$$

$$\sigma_\Delta^2 = \sigma_{H_t}^2 + \sigma_{H_{t+\Delta t}}^2 \tag{9}$$

Assuming that the errors of observations are distributed normally $\varepsilon \sim N(0, \sigma^2)$, then the adjustment parameters dH_t and $dH_{t+\Delta t}$ would also be distributed normally. Since the displacements Δ are differences between dH_t and $dH_{t+\Delta t}$, under the assumption of an equal approximate initial height dH^0 of the random benchmark in both epoch adjustments, the displacements Δ would also be normally distributed. Thus, the ratio (10) would follow the standard Normal distribution.

$$T = \frac{\Delta}{\sigma_\Delta} \sim N(0,1) \tag{10}$$

If $T \geq N(0, 1, \alpha)$, then we can conclude at a significance level higher than $1-\propto$ that the quantity Δ is a real displacement, but not a measurement error. More sophisticated estimation of the statistic T through simulations is also possible (Savšek-Safić et al., 2006).

Finally, the velocity of a random nodal benchmark common to both epochs can be computed using equation (11).

$$V = \frac{\Delta}{\Delta t} = \frac{\Delta}{23.22} \tag{11}$$

The accuracy of the speeds *V* is directly given by (12) (Kääriäinen, 1966, pp. 59).

$$\sigma_V = \frac{\sigma_\Delta}{\Delta t} = \frac{\sigma_\Delta}{23.22} \tag{12}$$

Thus, statistic (10) is a quite valid estimator for assessing the significance of the speeds in equation (11).

$$T = \frac{V}{\sigma_V} = \frac{\Delta/\Delta t}{\sigma_\Delta/\Delta t} = \frac{\Delta}{\sigma_\Delta} \sim N(0,1) \tag{13}$$

3. Results

Table 1 shows the adjusted heights of common nodal benchmarks for epochs 1955.22 and 1978.44, along with their standard deviations. Table 4 presents the F-Test results comparing the a posteriori accuracy from both adjustments. Table 3 lists the estimated benchmark displacements, their velocities, associated uncertainties, and statistical significance. Figure 8 maps the recent ground movement speeds across Bulgaria, based on this analysis.

Table 1. Adjustment results.

Benchmark	1955.22		1978.44	
	H m	σ_H (mm)	H m	σ_H (mm)
Burgas	82.46760	2.06	82.45194	2.25
G. Alexandrovo	154.12936	2.09	154.09012	2.26
Elhovo	128.47842	2.09	128.43560	2.27
Tutrakan	35.20201	4.49	35.22153	4.61
Kazanlak	376.47839	2.60	376.43639	2.79
Levski	70.47135	3.19	70.42584	3.24
Momchilgrad	250.81428	4.42	250.85406	4.35
Dospat	1196.32312	4.57	1196.36630	4.35
Zlatitsa	674.71126	3.01	674.64775	3.02
Knezha	161.39332	3.14	161.35912	3.18
Vidin	42.06641	4.69	42.03004	5.30
Sofia (N. Iskar)	515.92643	3.04	515.87861	3.07
Petrich	110.40077	5.03	110.39552	4.54

Table 2. Test of the a posteriori variance factors for both epochs for compatibility, performed at the significance level $\alpha = 0.05$.

Statistic	**Value**
$\sigma_{1955.22}$	2.8166
$\sigma_{1978.44}$	2.2354
Fobs	1.5876
F(7,7,0.05)	3.7870
p-Value	0.2784

Table 3. Estimated benchmark displacements, their speed, accuracy, and statistical significance.

Benchmark	Δ (mm)	V (mm/year)	σ_v (mm/year)	Tobs.	N(0, 1, T_{obs}, two-tail)
Burgas	-15.65	-0.67	0.13	5.13	**2.862E-07**
G. Alexandrovo	-39.25	-1.69	0.13	12.75	**0.000E+00**
Elhovo	-42.82	-1.84	0.13	13.86	**0.000E+00**
Tutrakan	19.52	0.84	0.28	3.03	**2.406E-03**
Kazanlak	-42.00	-1.81	0.16	11.01	**0.000E+00**
Levski	-45.51	-1.96	0.20	10.00	**0.000E+00**
Momchilgrad	39.78	1.71	0.27	6.41	**1.433E-10**
Dospat	43.18	1.86	0.27	6.85	**7.556E-12**
Zlatitsa	-63.52	-2.74	0.18	14.91	**0.000E+00**
Knezha	-34.20	-1.47	0.19	7.65	**1.954E-14**
Vidin	-36.37	-1.57	0.30	5.14	**2.749E-07**
Sofia (N. Iskar)	-47.82	-2.06	0.19	11.07	**0.000E+00**
Petrich	-5.25	-0.23	0.29	0.77	4.385E-01

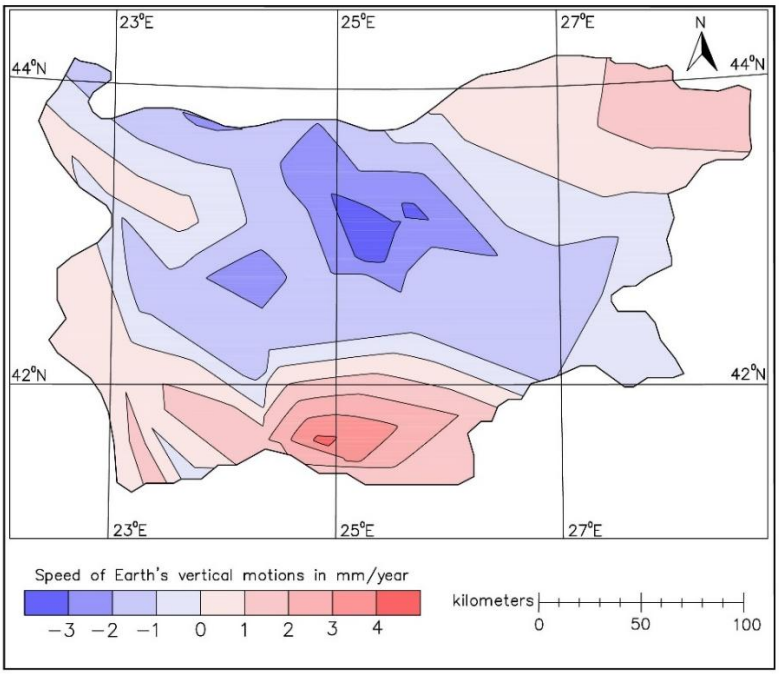


Fig.8. The new map of the recent vertical movements of the Earth's crust in Bulgaria.

4. Discussion

Table 1 shows that the standard deviations of the adjusted nodal benchmark heights are within 5 mm for both network epochs, indicating strong consistency. The adjusted benchmark heights are highly similar between epochs, a fact confirmed statistically by the compatibility test of a posteriori variance factors in Table 3. These results show that both networks have equal accuracy, meaning the estimation of vertical displacements won't be compromised by poor weighting models, outliers, or major differences in network design.

As a result, the displacement estimates, along with their velocities and accuracy figures in Table 2, are statistically valid. Figure 8, based on Table 2, illustrates that Bulgaria can be divided into five regions according to the direction and magnitude of vertical crustal movement. The first one is the Eastern part of the Lodogorie-Dobrudzha Swell. This part of Bulgaria is rising by 1 mm/year. The boundary, that is to say, the velocity zero line, starts from Varna Bay, passes along the rivers Pomoriyska, Beli Lom, and Rusenski Lom. Thus, this zero line is very close to the Intramoesian Fault sketched by Stanciu and Ioane (2023, Figure 2) and coincides with a group of active faults given in this sketch.

The second part includes the rest of the Moesian Platform (Lom Depression, Iskar-Yantra step, South Moesian Platform margin), the Forebalkan, Central and East Balkan, Srednogorie unit, and the Strandzha-Sakar massifs. This entire territory is sinking at a rate of 1-2 mm/year. The only exception is the area along the Yantra river between the towns of Gabrovo and Gorna Oriyahovitsa, where the velocities of sinking are up to -3.5 mm/year. This sinking is likely a logical result of the intensive seismic activities in this place (Stanciu and Ioane, 2023, Figure 11). Our results confirm the subsidence of the Lom depression at a rate of 1.5-2.5 mm/year.

The south boundary of the second part is the Maritsa fault. South of this line, which is an agglomerate of important neotectonic faults, lies the third tectonic part in Bulgaria, the Morava-Rhodope zone, which is rising by 1-2 mm/year. The supreme rise of +4.3mm/year we registered in the area of Smolyan town, which is close to the highest Rhodopes' peak – Golyam Perelik. Other rises more significant than 3 mm/year we yielded in Madan / 3.5 mm/year /, and Zlatograd / 3.2 mm/year /. Considering the Pirin Mountain, we found that the territory of Papas Chayr passage is rising by 1.7 mm/year. The area around Kroupnik town, known for its extensive seismic activity, is growing by 1.4 mm per year.

The only zone south of the Maritsa fault, where we registered sinking, is the area of the Mesta River fault. According to our calculations, the area around Gotse Delchev is sinking by 0.5 mm/year.

The last fifth zone is the West Balkan unit, located west of the Iskar River. The Forebalkan fault in the North surrounds this zone, the Iskar River fault in the East, and the boundary between the Forebalkan and the Balkan in the South. The registered rise between Belogradchik and the Prevala passage is approximately 0.8 mm/year, which is higher than the standard errors of the velocity multiplied by three.

In conclusion, the velocities of vertical crustal movements in Bulgaria generally range within ±3 mm/year, except in the area surrounding the peak of Golyam Perelik in the Central Rhodopes, where they are over +4mm/year. These rates are comparable to those reported for Lithuania (Puziene et al., 2013). According to Puziene (2011), vertical movement velocities of this magnitude can significantly compromise the stability and accuracy of national reference systems. An analysis of selected lines in the Lithuanian levelling network demonstrated that vertical tectonic motions of similar rates reduce the accuracy of first-order levelling by approximately 67% over a period of ten years.

Complete degradation of the highest-order precise levelling network is expected within about 20 years, while second-order levelling networks may remain reliable for up to 30 years. Based on these findings, the highest-order Bulgarian levelling network should be remeasured at intervals of approximately 15–20 years to account for ongoing vertical tectonic activity in the region.

Another important aspect of the derived vertical tectonic velocities concerns the relationship between their signs and magnitudes and seismic hazard assessments. As discussed above, the boundaries between zones of positive and negative vertical velocities delineated on the new vertical tectonic map of Bulgaria closely follow well-known active faults. Figure 9 illustrates the spatial relationship between the epicentres of the strongest earthquakes in Bulgaria over the past two centuries and the boundaries separating zones of opposite vertical motion. Notably, the epicentres of major earthquakes, such as those in Kresna (1904, surface-wave magnitude Ms > 7) (Bayliss and Burton, 2007; Dimitrov and Nakov, 2022), Plovdiv and Chirpan (1928, Ms > 7) (Bayliss and Burton, 2007), and Shabla (1901, Ms = 7.2) (Matova, 2000), are located along these boundaries. These zones coincide with Bulgaria's most significant active fault systems, including the Maritsa Fault (as depicted by Dabovski et al., 2002, Fig. 1), the Struma Fault, and the Intramoesian Fault (Stanciu and Ioane, 2023, Fig. 2).

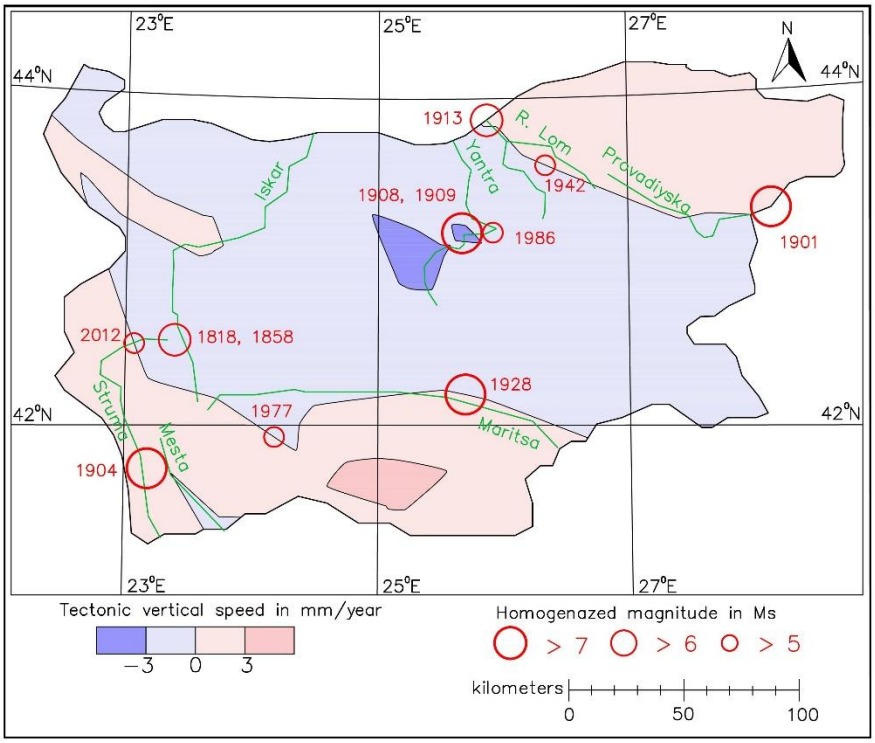


Fig. 9. The largest earthquake in Bulgaria in the last two centuries.

As shown in Figure 9, the epicentres of the earthquakes in Gorna Oryahovitsa (1908 and 1909, with surface-wave magnitudes M >7) (Bayliss and Burton, 2007) and in

Stazhitsa (1986, Ms ≈ 5.6) are situated within the zone exhibiting the highest subsidence in Bulgaria. The epicentres of the earthquakes in Razgrad (1942, Ms ≈ 5.1), Velingrad (1977, Ms ≈ 5.2), the Sofia region (1818 and 1858, Ms ≈ 5–7) (Dimitrov and Nakov, 2022), and Pernik (1977, $M_S \approx 5.6$) (Dimitrov and Nakov, 2022; Solakov et al., 2020) also show strong spatial correspondence with the identified zones. Furthermore, the majority of earthquakes recorded in Bulgaria between 1981 and 2019 (Solakov et al., 2020) are concentrated within these areas. This demonstrates a strong correlation between the boundaries separating zones of positive and negative vertical tectonic motions and seismic hazard. Consequently, periodic remeasurement of the national levelling network is necessary to monitor tectonic activity and to mitigate the risk of damage to buildings and infrastructure, as well as loss of life, as observed during the largest earthquakes in Bulgaria (Bayliss and Burton, 2007; Matova, 2000). In addition, Figure 9 shows a strong correspondence between the zone boundaries and tectonic depressions, the major river flows in Bulgaria, and the epicentres of the country's largest earthquakes. The estimated distances between them are generally less than 10 km. In more cases, even below 3 km – Plovdiv and Chirpan /Maritsa river/, Kresna and Pernic /Struma river/, Gorna Oryahovitsa, Veliko Tarnovo and Stazhitsa /Yantra river/, Sofia /Iskar river/. These differences can be attributed to the generation of tectonic isolines, residual calculation uncertainties, and seismic and levelling measurement errors.

5. Conclusion

This paper introduces a new approach for estimating vertical displacements of the Earth's crust and their rates. The key steps in the procedure include:

- Selecting levelling network values that minimise loop closure errors through separate adjustments for each epoch.
- Reducing the influence of network geometry on final results using IDW.
- Analysing the consistency between the two epoch networks.
- Estimating the vertical displacements and displacement rates of benchmarks.

The result is a new, more accurate and reliable map of recent vertical crustal movements in Bulgaria, which closely aligns with known tectonically active faults. According to the findings, the Morava-Rhodope zone, the Western Balkan Unit, and the Dobrudzha Arch are uplifting, while the Srednogorie Unit, the Moesian Platform (excluding the Dobrudzha Arch), and the Balkan Unit (excluding the West Balkan) are subsiding. Vertical displacement rates range from -3.5 mm/year to +4.0 mm/year, indicating notable tectonic activity. The ratios between the determined Earth's crust annual speeds and their standard deviations, except the benchmarks in Petrich, are in the range between 3.25 and 15.69. In addition, a strong correlation was identified between the boundaries separating zones of positive and negative vertical tectonic motions and the epicentres of the largest earthquakes in Bulgaria. Consequently, these dynamics must be considered when adjusting or establishing national geodetic reference systems, not only to ensure an accurate height reference system but, more importantly, to help mitigate potential loss of life and damage associated with earthquakes.

Because of the fast evaluation of Global Navigation Satellite Systems (GNSS) accuracy (Balodis et al., 2019; Borowski et al., 2025; Celms et al., 2024; Haritonova, 2019), our further research would be dedicated to the concordance between the Earth's vertical motions in Bulgaria obtained by precise levelling and GNSS technologies.

Acknowledgements

The authors express gratitude to the Centre for Scientific Research and Design at the UACEG for funding and support for the completed research project under contract No BN315/2025 on the topic "Optimisation of levelling networks by studying the applicability of resampling methods."

References:

Balodis, J., Varna, I., Haritonova, D., Morozova, K. (2019). Coordinate Analysis of Latvian CORS Stations. *Baltic Journal of Modern Computing*, **7(**4), 513-524. https://doi.org/10.22364/bjmc.2019.7.4.05

Bayliss, T. J., Burton, P. W. (2007). A new earthquake catalogue for Bulgaria and the conterminous Balkan high hazard region, *Nat. Hazards Earth Syst. Sci*., 7, 345–359, https://doi.org/10.5194/nhess-7-345-2007

Belyashki, T. (2012). A new map of the recent vertical movements of the Earth's crust in the territory of Bulgaria. *Geodezia, kartographia, zemeustroistvo*. 3-4:3-5 (in Bulgarian).

Borowski, L., Banasik, P., Maciuk, K. (2025). Application of GNSS-levelling for updating the base vertical network, *Journal of Applied Geodesy*, DOI: 10.1515/jag-2024-0096.

Caspary, W. F. (2000). Concepts of Network and Deformation Analysis, Monograph 11, Third Edition, School of Geomatic Engineering, The University of New South Wales, Australia, ISBN 0-85839-044-2.

Celms, A., Grinberga, L., Lidumnieks, T., Luksa, J., Brinkmanis-Brimanis, M. (2024). Assessment of accuracy of GNSS measurement models using base station and radio module solutions. *Baltic Journal of Modern Computing*, 12(4):454–462. https://doi.org/10.22364/bjmc.2024.12.4.07

Cvetkov, V. (2024). Averages, entropy and observations, *Deutsche Internationale Zeitschrift für zeitgenössische Wissenschaft*, 93:52–7. https://doi.org/10.5281/zenodo.14289165

Cvetkov, V. (2024). Entropy, observations and some standard deviations, *Deutsche Internationale Zeitschrift für zeitgenössische Wissenschaft*, 94:13–8. https://doi.org/10.5281/zenodo.14531534

Cvetkov, V. (2024). On initial data in adjustments of the geometric levelling networks (on the mean of paired observations). *Journal of Geodetic Science*, 14(1):20220170. https://doi.org/10.1515/jogs-2022-0170.

Cvetkov, V. (2024). Geometric Levelling Data and Some Systematic Faults in Their Treatment, *Inżynieria Mineralna - Journal of the Polish Mineral Engineering Society*, 2024, 1(1), 293–300. https://doi.org/10.29227/IM-2024-01-32

Cvetkov, V. (2023). Two adjustments of the second levelling of Finland by using nonconventional weights, *Journal of Geodetic Science*, 13(1), https://doi.org/10.1515/jogs-2022-0148.

Cvetkov, V. (2022). Discrepancies and misclosures in the Third Levelling of Bulgaria, *Annual of the University of Architecture, Civil Engineering and Geodesy, Sofia*, 55(1):195-201. https://old.uacg.bg/UserFiles/File/UACEG_Annual/2022/%D0%91%D1%80%D0%BE%D0%B9%201/14-G-3.pdf (accessed on 10.01.2026)

Cvetkov, V., Gospodinov, S. (2023). Inverse Absolute Height Weighting in the Highest Order Levelling Networks, In *EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023*, EGU23-4219, https://doi.org/10.5194/egusphere-egu23-4219.

Dabovski C., Boyanov, I., Khrischev, K., Nikolov, T., Sapounov, I., Yanev, Y., Dimitrov, N., Nakov, R. (2022). GPS Results from Long Time Monitoring of Geodynamic Processes in South-Western Bulgaria. *Appl. Sci.* 2022, 12, 2682. https://doi.org/10.3390/app12052682

Gospodinov, S., Peneva, E., Penev, P. (2022). A specific approach to least squares adjustment of the state levelling network, *Proceedings of the 22nd International Multidisciplinary Scientific GeoConference SGEM 2022*, 22(2.1):163-169. https://doi.org/10.5593/sgem2022/2.1/s09.20

Hamza, V., Stopar, B., Ambrožič T. (2020). Deformation Analysis: the Caspary Approach. *Geodetski vestnik*, 64 (1), 68-88. https://doi.org/10.15292/geodetski-vestnik.2020.01.68-88

Haritonova, D. (2019). The impact of the baltic sea non-Tidal loading on gnss station coordinate time series: The case of Latvia. *Baltic Journal of Modern Computing*, 7(4), 541-549. https://doi.org/10.22364/bjmc.2019.7.4.07

Kanev, D., Mladenovski, M. (1969). The recent vertical movements of the Earth's crust in Bulgaria. *Journal of the Bulgarian Geographical Society*. IX:17–26 (in Bulgarian).

Kääriäinen, E. (1966). The Second Levelling of Finland in 1935–1955, *Publications of the Finnish Geodetic Institute No. 61*, Helsinki. http://hdl.handle.net/10138/347072

Lyszkowicz, A., Leonczyk, M. (2006) The Fourth Precise Levelling Campaign of Poland in 1999–2003, Promoting Land Administration and Good Governance, 5th FIG Regional Conference Accra, Ghana, 8-11 March 2006.

Matova, M. (2000). Recent geological activity along the northeast Bulgarian Black Sea coast, *Gcological Quarterly*, 44 (4): 355-361. https://gq.pgi.gov.pl/article/view/8109

Niemeier, W. (1981). Statistical Tests for Detecting Movements in Repeatedly Measured Geodetic Networks,Editor(s): P. Vyskočil, R. Green, H. Mälzer, Developments in Geotectonics, Elsevier, 16:335-351, ISSN 0419-0254, ISBN 9780444419538, https://doi.org/10.1016/B978-0-444-41953-8.50041-4.

Puziene, R.(2013). Impact of vertical ground movements on accuracy of reference geodetic levelling networks. *Selected papers of the 8th International Conference May 19-20, 2011, Vilnius, Lithuania*, pp. 1440–1447.

Puziene, R., Anikeninene, A., Stanionis, A. (2013). A model of recent vertical movements of Earth's surface of the territory of Lithuania. *Advanced Materials Research*, 804:279–286. https://doi.org/10.4028/www.scientific.net/AMR.804.279

Sacher, M., Belyashki, T., Liebsch, G., Ihde, J. (2004). Status of the UELN/EVRS database and the last UELN results. In Proceedings of the EUREF Symposium 2004. EUREF. https://www.euref.eu/sites/default/files/symposia/book2004/5-1.pdf

Savšek-Safić, S., Ambrožič, T., Stopar, B., Turk, G. (2006). Determination of point displacements in the geodetic network. *Journal of Surveying Engineering*, 132(2), 58-63. https://doi.org/10.1061/(asce)0733-9453(2006)132:2(58).

Setan, H., Singh, R. (2001). Deformation Analysis of a Geodetic Monitoring Network, *Geomatica*, 55(3):333-346, ISSN 1195-1036, https://doi.org/10.5623/geomat-2001-0039.

Solakov, D., Simeonova, S., Raykova, P., Aleksandrova, I. (2020). *Catalogue of the earthquakes in Bulgaria and surroundings since 1981 [Data set]*. National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences. https://doi.org/10.34975/CTLG-2020.V.1

Spiridonov, H., Georgiev, N. (2003), Study of the Neotectonics and Geodynamics of the Republic of Bulgaria, *Aerospace Research in Bulgaria*, 17:84-96.

Stanciu, I., Ioane, D. (2023). Geomorphological and Neotectonic Structures Studied in the Southern Part of the Moesian Platform in Romania. *Geographies*, 3:743–762, https://doi.org/10.3390/geographies3040040.

Zagorchev, I. (2002). Structure and Alpine evolution of Bulgaria, Geologica Balcanica, 32(2-4):9-15. https://doi.org/10.52321/GeolBalc.32.2-4.9

Zagorchev, I. (2009). Geomorphological zonation of Bulgaria. Principles and state of the art, *Proceedings of the Bulgarian Academy of Sciences*, 628(8):981-992. https://www.researchgate.net/publication/271190963_Geomorphological_zonation_of_Bulgaria_Principles_and_state_of_the_art

Appendix A

Pseudocode of 3^n Selection Procedure

```
INTEGER n, k
INTEGER ARRAY loops
STRING MATRIX Benchmarks
STRING MATRIX Weights
DOUBLE MATRIX Elevations
DOUBLE min_Mu ← +∞
DOUBLE ARRAY best_Elevations
DOUBLE ARRAY Adj_Elevations
STRING Benchmark_RMSE

FUNCTION Constructor(benchmarks, weights, elevations)
    Benchmarks ← benchmarks
    Weights ← weights
    Elevations ← elevations
    n ← number of rows in elevations
    k ← number of columns in elevations
    loops ← new INTEGER[n]
    best_Elevations ← new DOUBLE[n]
    Adj_Elevations ← new DOUBLE[n]
    CALL Calculate(Benchmarks, Weights, Elevations)
END FUNCTION


FUNCTION Adjusted_Elevations()
    STRING result ← empty
    FOR i from 0 to n-1
        append Adj_Elevations[i] formatted to 5 decimals to
result
    END FOR
    RETURN result
END FUNCTION


FUNCTION Calculate(Benchmarks, Weights, Elevations)
    CALL GenerateCombinations(0)
END FUNCTION

/*
This generates all possible combinations (k^n).
WeightedLevelling(Weights, Benchmarks) performs the
geometric levelling adjustment.
*/
```

```
FUNCTION GenerateCombinations(current)
    IF current = n THEN
        adj ← new WeightedLevelling(Weights, Benchmarks)
        FOR i from 0 to n-1
            adj.Elevations[i,0] ← Elevations[i, loops[i]-1]
        END FOR
        CALL adj.Calculate()

        IF adj.Mu < min_Mu THEN
            min_Mu ← adj.Mu
            FOR i from 0 to n-1
                best_Elevations[i] ← adj.Elevations[i,0]
                Adj_Elevations[i] ← adj.Adj_Elevations[i,0]
            END FOR

            Benchmark_RMSE ← adj.BenchMark_RMSE()
            BUILD output summary string
            (timestamp, Mu, best_Elevations,
adj.Elevations, Benchmark_RMSE)
        END IF

        RETURN
    END IF

    FOR counter from 1 to k
        loops[current] ← counter
        CALL GenerateCombinations(current + 1)
    END FOR
END FUNCTION
```

Appendix B

The Second Levelling of Bulgaria data that were used in the study

From	To	Length (km)	Height differences in (m)		
			I	II	Mean
Knezha	Vidin	163.168	-119.33810	**-119.32692**	-119.33251
Zlatitsa	Knezha	117.062	-513.28172	**-513.31777**	-513.29975
Sofia	Zlatitsa	93.805	158.75188	**158.78483**	158.76836
Vidin	Sofia	255.104	473.83649	473.88345	**473.85997**
Knezha	Levski	104.418	**-90.92189**	-90.91337	-90.91763
Levski	Kazanlak	174.666	305.97630	**306.00681**	305.99156
Zlatitsa	Kazanlak	127.928	-298.25715	**-298.23291**	-298.24503
Levski	Tutrakan	172.996	**-35.26736**	-35.21556	-35.24146
Tutrakan	Varna	211.602	**30.17987**	30.18738	30.18362
Varna	Burgas	136.621	**17.09283**	17.12000	17.10642
Burgas	G. Alexandrovo	71.044	**71.66161**	71.65542	71.65852
G. Alexandrovo	Kazanlak	133.286	**222.34934**	222.38840	222.36887
Elhovo	Burgas	114.001	**-46.01353**	-46.02272	-46.01813
Elhovo	G. Alexandrovo	69.470	25.64586	**25.65108**	25.64847
Momchilgrad	Elhovo	253.155	-122.37364	-122.30770	**-122.34067**
Momchilgrad	Kazanlak	171.824	125.63275	**125.66310**	125.64792
Momchilgrad	Dospat	201.120	**945.51263**	945.51825	945.51544
Dospat	Zlatitsa	177.768	**-521.60987**	-521.60246	-521.60617
Dospat	Petrich	146.882	-1085.97929	**-1085.92241**	-1085.95085
Petrich	Sofia	260.335	405.47898	**405.52389**	405.50144

Note: The bolded values are the height difference values selected by 3^{20} adjustments. They produce adjustment with the least aposteory standard error.

Appendix C

The Third Levelling of Bulgaria data that were used in the study

From	To	Length (km)	Height differences in (m)		
			I	II	Mean
Knezha	Vidin	185.971	-119.30612	**-119.32912**	-119.31762
Zlatitsa	Knezha	126.510	**-513.28901**	-513.32191	-513.30546
Sofia	Zlatitsa	95.510	158.78266	**158.76916**	158.77591
Vidin	Sofia	406.478	**473.84495**	473.81374	473.82935
Knezha	Levski	112.309	**-90.93347**	-90.93774	-90.93561
Levski	Kazanlak	184.355	**306.00943**	305.95973	305.98458
Zlatitsa	Kazanlak	126.260	**-298.21085**	-298.20101	-298.20593
Levski	Tutrakan	191.096	**-35.20747**	-35.21508	-35.21128
Tutrakan	Varna	225.754	**30.14512**	30.11246	30.12879
Varna	Burgas	151.930	17.06735	**17.07590**	17.07162
Burgas	G. Alexandrovo	71.505	71.65568	71.62065	**71.63817**
G. Alexandrovo	Kazanlak	139.409	222.36740	222.32509	**222.34625**
Elhovo	Burgas	307.110	-45.92758	**-45.96507**	-45.94632
Elhovo	G. Alexandrovo	66.459	**25.65452**	25.65348	25.65400
Momchilgrad	Elhovo	261.637	**-122.39892**	-122.38961	-122.39427
Momchilgrad	Kazanlak	184.636	**125.57867**	125.56773	125.57320
Momchilgrad	Dospat	203.882	945.55511	**945.51450**	945.53481
Dospat	Zlatitsa	190.734	-521.67446	**-521.71862**	-521.69654
Dospat	Petrich	147.738	-1085.96916	**-1085.97044**	-1085.96980
Petrich	Sofia	204.984	**405.48552**	405.45665	405.47108

Note: The bolded values are the height difference values selected by 3^{20} adjustments. They produce adjustment with the least aposteory standard error.

Appendix D

Adjusted benchmark heights in the Second Levelling of Bulgaria after 3^n selections and IDW(p), where p = {4, 5, 6, 7, 8}

Benchmark	p = 4	p = 5	p = 6	p = 7	p = 8
Burgas	82.46714	82.46740	82.46760	82.46775	82.46785
G. Alexandrovo	154.12905	154.12923	154.12936	154.12946	154.12953
Elhovo	128.47827	128.47836	128.47842	128.47846	128.47850
Tutrakan	35.20078	35.20142	35.20201	35.20253	35.20297
Kazanlak	376.47792	376.47817	376.47839	376.47857	376.47871
Levski	70.47055	70.47098	70.47135	70.47164	70.47185
Momchilgrad	250.81400	250.81412	250.81428	250.81446	250.81463
Dospat	1196.32312	1196.32311	1196.32312	1196.32313	1196.32311
Zlatitsa	674.71084	674.71105	674.71126	674.71145	674.71160
Knezha	161.39269	161.39302	161.39332	161.39357	161.39376
Vidin	42.06580	42.06612	42.06641	42.06665	42.06684
Sofia (N. Iskar)	515.92599	515.92621	515.92643	515.92662	515.92677
Petrich	110.40084	110.40079	110.40077	110.40075	110.40073

Adjusted benchmark heights in the Third Levelling of Bulgaria after 3^n selections and IDW(p), where p = {4, 5, 6, 7, 8}

Benchmark	p = 4	p = 5	p = 6	p = 7	p = 8
Burgas	82.45258	82.45222	82.45194	82.45173	82.45157
G. Alexandrovo	154.09076	154.09040	154.09012	154.08990	154.08973
Elhovo	128.43627	128.43589	128.43560	128.43538	128.43521
Tutrakan	35.22314	35.22234	35.22153	35.22077	35.22010
Kazanlak	376.43660	376.43654	376.43639	376.43622	376.43606
Levski	70.42703	70.42643	70.42584	70.42533	70.42493
Momchilgrad	250.85287	250.85357	250.85406	250.85442	250.85469
Dospat	1196.36637	1196.36640	1196.36630	1196.36616	1196.36602
Zlatitsa	674.64831	674.64807	674.64775	674.64743	674.64715
Knezha	161.36005	161.35960	161.35912	161.35869	161.35833
Vidin	42.03107	42.03055	42.03004	42.02958	42.02921
Sofia (N. Iskar)	515.87922	515.87895	515.87861	515.87828	515.87800
Petrich	110.39546	110.39555	110.39552	110.39545	110.39537

Estimated vertical speeds in mm/year of the nodat benchmarks under different values of the power parameter p, applied in the IDW step, where p = {4, 5, 6, 7, 8}

Benchmark	p = 4	p = 5	p = 6	p = 7	p = 8
Burgas	-0.63	-0.65	-0.67	-0.69	-0.70
G. Alexandrovo	-1.65	-1.67	-1.69	-1.70	-1.71
Elhovo	-1.81	-1.83	-1.84	-1.86	-1.86
Tutrakan	0.96	0.90	0.84	0.79	0.74
Kazanlak	-1.78	-1.79	-1.81	-1.82	-1.84
Levski	-1.87	-1.92	-1.96	-1.99	-2.02
Momchilgrad	1.67	1.70	1.71	1.72	1.73
Dospat	1.86	1.86	1.86	1.85	1.85
Zlatitsa	-2.69	-2.71	-2.74	-2.76	-2.78
Knezha	-1.41	-1.44	-1.47	-1.50	-1.53
Vidin	-1.50	-1.53	-1.57	-1.60	-1.62
Sofia (N. Iskar)	-2.01	-2.04	-2.06	-2.08	-2.10
Petrich	-0.23	-0.23	-0.23	-0.23	-0.23

Received December 2, 2025, revised March 7, 2026, accepted March 10, 2026