Biljana Risteska Stojkoska • Smilka Janeska Sarkanjac (Eds.)

16TH ICT INNOVATIONS CONFERENCE

TechConvergence: AI, Business, and Startup Synergy

WEB PROCEEDINGS
28-30 September 2024
Metropol Lake Resort, Ohrid, Macedonia
ICT ACT, 2024
Online edition published on http://ictinnovations.org

Biljana Risteska Stojkoska • Smilka Janeska Sarkanjac (Eds.)

16TH ICT INNOVATIONS CONFERENCE,

TechConvergence: AI, Business, and Startup Synergy

WEB PROCEEDINGS

28-30 September 2024

Metropol Lake Resort, Ohrid, Macedonia

ISBN 978-608-65468-4-7 © ICT ACT

Publisher: Society of Information and Communication Technologies (ICT-ACT)

Online edition published on http://ictinnovations.org

Skopje, January 2025

Edited by: Biljana Risteska Stojkoska, Smilka Janeska Sarkanjac

Technical support: Ilinka Ivanoska, Ana Todorovska

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

004:621.39(062)

ICT innovations conference (16; 2024; Ohrid)
16th ICT innovations conference [Електронски извор] : techconvergence: AI, business, and startup synergy : web proceedingS : 28-30 September 2024 Metropol lake resort, Ohrid, Macedonia ICT ACT, 2024 / (eds.) Biljana Risteska Stojkoska ї Smilka Janeska Sarkanjac. - Skopje : Society of information and communication technologies ICT-ACT, 2025

Начин на пристапување (URL): https://ictinnovations.org/wp-content/uploads/2025/01/webproc_final.pdf. - Текст во PDF формат, содржи 308 стр., илустр. - Наслов преземен од екранот. - Опис на изворот на ден 24.01.2025. - Библиографија кон трудовите

ISBN 978-608-65468-4-7

а) Информациско-комуникациски технологии -- Примена -- Собири

COBISS.MK-ID 65144837

Preface

We are proud to present the proceedings of the 16th International Conference ICT Innovations 2024, held in Ohrid, North Macedonia, from September 28-30, 2024. This year's conference marks another significant milestone in our ongoing commitment to fostering groundbreaking research and innovation in information and communication technologies (ICT). Under the theme, "TechConvergence: AI, Business, and Startup Synergy," the conference brought together over 120 researchers, practitioners, and industry leaders from across the globe to explore the synergies between AI, business intelligence, and entrepreneurship.

The ICT Innovations conference series, organized by the Macedonian Society of Information and Communication Technologies (ICT-ACT) and supported by the Faculty of Computer Science and Engineering (FCSE) in Skopje, built a distinguished reputation as a platform for presenting both fundamental and applied research. Over the years, this conference became a critical venue for sharing innovative solutions and scientific discoveries, continually addressing the most pressing challenges and opportunities within the field of ICT.

In this 2024 edition, we turned our focus to how data science intersected with the business and start-up world, encouraging interdisciplinary collaboration and knowledge-sharing. As technology evolved rapidly, these collaborations became essential for harnessing the transformative potential of innovations. The central focus of this event revolved around the convergence of data science and business strategies, with particular emphasis on how these technologies shaped the future of entrepreneurship.

This volume contained 21 full papers (plus 21 short papers in the Web Proceedings edition), which were carefully reviewed and selected from 80 high-quality submissions. These papers covered a wide range of topics, including machine learning, network science, digital transformation, natural language processing, and more. The review process was rigorous, with about 100 reviewers from 35 countries providing detailed feedback. Each submission was evaluated by at least three experts in the field, ensuring that the selected papers met the high standards of academic excellence and originality that this conference is known for.

The program featured two distinguished keynote speakers who demonstrated the convergence of technology, innovation, and business growth. Their presentations showcased how advancements in ICT transformed industries and drove entrepreneurial success. Paul Kayne discussed his work at Palatin Technologies, exploring how ICT and big data were used to understand genomes and the melanocortin system, presenting new approaches to treating inflammatory diseases. He also proposed ways to share genomics data while safeguarding intellectual property. Dejan Zvekic, a key figure in the regional IT start-up scene, shared his entrepreneurial journey, detailing how he transformed the fashion industry at Material Exchange, expanded Plugin76, and successfully integrated it

into PTC Inc. His presentation highlighted the importance of strategic insight, market awareness, and leveraging opportunities for growth.

Alongside the main conference, participants had the opportunity to engage in seven specialized workshops focused on entrepreneurship, human-machine collaboration, and data analytics for business intelligence: Innovations in anti-drone technologies; Black-box and explainable artificial intelligence methods; Blended research on air pollution; Interactive data science; CyberMACS; and CHATMED.

The workshop titled "Women in STEM", where a diverse group of speakers, including Aneta Antova Pesheva, Eliot Bytyci, and Afrodita Shalja, addressed the critical issue of underrepresentation of women in Informatics across all educational and professional levels. The speakers highlighted the slow progress in increasing female participation in STEM fields, despite ongoing efforts in Europe. They encouraged researchers to submit papers on initiatives aimed at engaging and retaining female students and professionals, fostering a supportive environment for women pursuing careers in these disciplines.

All these workshops provided a hands-on experience, allowing researchers and practitioners to collaborate and explore new ideas in an interactive setting. The conference also offered a variety of social events aimed at fostering connections among participants, a tradition that has been highly valued since the conference's inception.

We extended our heartfelt thanks to all the authors who contributed their work to this year's proceedings, to the reviewers who ensured a fair and thorough evaluation process, and to all the participants who enriched the conference with their knowledge and expertise. Special thanks went to our generous sponsors, companies Netcetera and Ultra Computing; also to the organizing committee and the technical support team at FCSE, whose dedication and hard work were instrumental in making this conference a success. We were also deeply grateful to Ilinka Ivanoska for her invaluable assistance throughout the organization of the event.

As we concluded this edition of ICT Innovations, we looked ahead with excitement to future conferences, where we would continue to explore the frontiers of ICT and foster innovation across industries and disciplines. We invite you to join us at the 17th ICT Innovations conference in 2025, where we would continue this journey of scientific discovery and collaboration.

Sincerely,

November 2024

ICT Innovations 2024 Conference Chairs,
Biljana Risteska Stojkoska and Smilka Janeska Sarkanjac

Organization

General Chairs

Biljana Risteska Stojkoska	Ss. Cyril and Methodius University in Skopje, MK
Smilka Janeska Sarkanjac	Ss. Cyril and Methodius University in Skopje, MK

Program Commitee Chairs

Biljana Risteska Stojkoska	Ss. Cyril and Methodius University in Skopje, MK
Smilka Janeska Sarkanjac	Ss. Cyril and Methodius University in Skopje, MK
Ilinka Ivanoska	Ss. Cyril and Methodius University in Skopje, MK

Program Committee

Aleksandar Bojchevski	University of Cologne, DE
Aleksandar Stojmenski	Ss. Cyril and Methodius University in Skopje, MK
Aleksandra Mileva	University Goce Delcev, MK
Alessandro Cantelli-Forti	Lab RaSS National Laboratory - CNIT, IT
Amelia Badica	University of Craiova, RO
Ana Madevska Bogdanova	Ss. Cyril and Methodius University in Skopje, MK
Andrea Kulakov	Ss. Cyril and Methodius University in Skopje, MK
Andrej Brodnik	University of Ljubljana, SI
Andreja Naumoski	Ss. Cyril and Methodius University in Skopje, MK
Antonio De Nicola	ENEA, IT
Antun Balaz	Institute of Physics Belgrade, RS
Arianit Kurti	Linnaeus University, SE
Betim Cico	EPOKA University, Tirana, AL
Biljana Risteska Stojkoska	Ss. Cyril and Methodius University in Skopje, MK
Biljana Mileva Boshkoska	Faculty of information studies, SI
Blagoj Ristevski	Faculty of Information and Communication Technologies Bitola, MK
Bojan Ilijoski	Ss. Cyril and Methodius University in Skopje, MK
Bojana Koteska	Ss. Cyril and Methodius University in Skopje, MK
Boris Delibašić	University of Belgrade - Faculty of Organizational Sciences, RS
Dejan Gjorgjevikj	Ss. Cyril and Methodius University in Skopje, MK
Dejan Spasov	Ss. Cyril and Methodius University in Skopje, MK
Dilip Patel	London South Bank University, UK
Dimitar Trajanov	Ss. Cyril and Methodius University in Skopje, MK
Edmond Jajaga	University for Business and Technology, XV
Eftim Zdravevski	Ss. Cyril and Methodius University in Skopje, MK

Matus Pleva	Technical University of Košice, SVK
Melanija Mitrović	University of Niš, RS
Mile Jovanov	Ss. Cyril and Methodius University in Skopje, MK
Milos Jovanovik	Ss. Cyril and Methodius University in Skopje, MK
Milos Stojanovic	Visoka tehnicka skola Nis, RS
Miroslav Mirchev	Ss. Cyril and Methodius University in Skopje, MK
Monika Simjanoska	Ss. Cyril and Methodius University in Skopje, MK
Natasha Ilievska	Ss. Cyril and Methodius University in Skopje, MK
Natasha Stojkovikj	University Goce Delcev, MK
Nevena Ackovska	Ss. Cyril and Methodius University in Skopje, MK
Novica Nosović	Faculty of Electrical Engineering, University of Sarajevo, BiH
Özge Büyükdağlı	International University of Sarajevo, BiH
Pance Ribarski	Ss. Cyril and Methodius University in Skopje, MK
Pece Mitrevski	University St. Kliment Ohridski, Faculty of ICT - Bitola, MK
Periklis Chatzimisios	International Hellenic University, GR
Petar Sokoloski	Ss. Cyril and Methodius University in Skopje, MK
Petre Lameski	Ss. Cyril and Methodius University in Skopje, MK
Riste Stojanov	Ss. Cyril and Methodius University in Skopje, MK
Rossitza Goleva	New Bulgarian University, BG
Sashko Ristov	University of Innsbruck, AT
Sasho Gramatikov	Ss. Cyril and Methodius University in Skopje, MK
Sergio Ilarri	University of Zaragoza, ES
Shuxiang Xu	University of Tasmania, AU
Simona Samardjiska	Radboud University, NL
Slobodan Kalajdziski	Ss. Cyril and Methodius University in Skopje, MK
Smilka Janeska Sarkanjac	Ss. Cyril and Methodius University in Skopje, MK
Snezana Savoska	Faculty of Information and Communication Technologies, Bitola, MK
Stanimir Stoyanov	University of Plovdiv "Paisii Hilendarski", BG
Suzana Loshkovska	Ss. Cyril and Methodius University in Skopje, MK
Tarik Namas	International University of Sarajevo, BiH
Ustijana Rechkoska-Shikoska	UIST - Ohrid, MK
Vacius Jusas	Kaunas University of Technology, LT
Verica Bakeva	Ss. Cyril and Methodius University in Skopje, MK
Vesna Dimitrievska Ristovska	Ss. Cyril and Methodius University in Skopje, MK
Vesna Dimitrova	Ss. Cyril and Methodius University in Skopje, MK
Vesna Dimitrievska	Silicon Austria Labs, Villach, AT
Vladimir Trajkovik	Ss. Cyril and Methodius University in Skopje, MK
Vladimír Siládi	Matej Bel University, SVK
Zlatko Varbanov	Veliko Tarnovo University, BG

Technical Committee

Ilinka Ivanoska	Ss. Cyril and Methodius University in Skopje, MK
Ana Todorovska	Ss. Cyril and Methodius University in Skopje, MK
Mila Dodevska	Ss. Cyril and Methodius University in Skopje, MK
Marija Taneska	Ss. Cyril and Methodius University in Skopje, MK
Zorica Karapancheva	Ss. Cyril and Methodius University in Skopje, MK
Stefan Andonov	Ss. Cyril and Methodius University in Skopje, MK

Table of Contents

Session 6

Session 7

NATO Workshop

Session 1

Cryptocurrencies portfolio optimization with an EFA solution

Krassimira Stoyanova[1[0000-0003-1508-7156]], Rabab Benotsmane[2[0000-0002-0440-1229]] and Vladislava Grigorova[1[0009-0003-3260-786X]]

[1] Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Sofia, Bulgaria
[2] Institute of Automation and Infocommunication
University of Miskolc
Miskolc, Hungary

Abstract. Portfolio optimization of cryptocurrencies using evolutionary algorithms is a relatively recent topic in the financial literature. To optimize a green investment portfolio, this study aims to explore the virtual prospects of cryptocurrencies, specifically Bitcoin, Ethereum, and others. A portfolio of cryptocurrencies with an enhanced Firefly algorithm (EFA) is solved when combined with the Firefly algorithm and tabu search. This study contributes to the existing literature by providing an assessment of the digital asset benefits that prominent cryptocurrencies can offer within a green portfolio context.

Keywords: Cryptocurrencies, Firefly algorithm, Tabu search, Portfolio optimization.

1 Introduction

In recent years, the fact that cryptocurrencies and blockchain technology are at the core of the fourth industrial revolution and have the potential to significantly affect a variety of economic and financial sectors has become increasingly apparent [1-2]. By the end of 2023, the cryptocurrency Bitcoin outperforms all major traditional assets, including stocks, bonds, gold, and oil, in terms of profitability. Despite the challenging macroeconomic conditions and issues inside the cryptocurrency business, its year-over-year growth surpasses 160 percent. Nvidia is the one exception to Bitcoin's underperformance, having surged by 241 percent since the start of the year. The global interest in digital transformation is growing, as various sectors and enterprises see the necessity to rely on digital tools and processes due to new advances and improved technology procedures [3]. Bitcoin offers a decentralized payment method that is not limited by geographical boundaries or the monetary constraints imposed by federal authorities. A recent study suggests that cryptocurrencies like Bitcoin and others are better classified as technology-based products and emergent asset classes than traditional currencies or securities [4].

The potential for portfolio diversification and optimization is drawing a diverse group of investors, individuals, industry participants, and professionals to explore the investment alternatives of this emerging asset class [5-8].

There are several approaches to solving a portfolio optimization problem, but the most famous one is the Markowitz optimization problem [9]. Nevertheless, if more constraints are introduced, the quadratic method becomes unsuitable for solving the Markowitz issue. In this circumstance, it is advisable to utilize meta-algorithms. Meta-heuristic algorithms refer to algorithms that are commonly inspired by nature and are used to solve nonlinear problems with constraints. The most crucial algorithms include the Genetic Algorithm, Ant Colony Algorithm, Particle Swarm Optimization Algorithm, and Firefly algorithm (FA). Meta-heuristic algorithms, unlike precision-solving techniques, are suitable for tackling large-scale issues and can produce good solutions within a reasonable timeframe. When utilizing precise methodologies or meta-heuristic algorithms to address a problem, it is important to take into account the problem's dimensions and organization [10].

The many hybrid optimizers have undergone significant transformations over the past decade, demonstrating the practicality and effectiveness of utilizing hybridization to develop high-performance optimizers [11]. The Firefly algorithm, introduced by Yang in 2007 [12-14], is a heuristic optimization technique that utilizes a population-based approach to solve combinatorial and nonlinear optimization problems. The FA is highly efficient in identifying solutions and facilitates straightforward implementation. Unlike the Particle Swarm Algorithm, the optimization process of FA does not entail looping in a locally optimal solution. Instead, it employs a direct randomized diversification of the search.

A limited selection of results from these experiments indicates the potential usefulness of Tabu search in many contexts. The inquiry into employee scheduling [15] addressed issues that required solving integer programming problems with formulations using between one and four million variables. It took 22-24 minutes to get solutions that were within 98% of an upper bound on optimality. The study of [16] examined the issue of identifying the coherence of probabilities that indicate whether certain sets of phrases are true. The research also explored the inclusion of probability intervals, conditional probabilities, and minimal changes needed to ensure satisfiability. By combining a Tabu search method with an exact 0-1 nonlinear programming technique to generate columns for a master linear program, we were able to successfully solve a problem with up to fourteen variables. This is a threefold increase in problem size compared to earlier solutions. The quadratic assignment study conducted in reference [17] achieved the most optimal solutions for all evaluated issues from the existing literature, while also needing a shorter amount of CPU time compared to earlier reports. The method also achieved superior solutions compared to the best-known solution for a classical benchmark problem [18]. Additionally, the method consistently produced solutions of equal or higher quality compared to solutions obtained through simulating annealing, as observed in the maximum satisfiability, graph coloring, and traveling salesman studies [19-22]. The study [23] employed a hybrid strategy that utilized the tabu search algorithm to generate the most efficient path by minimizing the applied torque. The shift towards a more environmentally friendly economy requires the creation of digital book apps [24], as well as digital forms of payment like cryptocurrency and others. To avert crises [25], it is imperative to implement new environmentally friendly changes that will effectively address and overcome them.

2 The Crypto problem formulation

The crypto assets S_1 , S_2, … S_n ($n \geq 2$) with random returns are considered. Let a set of $n \in$ N crypto assets be given. At time $t_0 \in$ R, each asset i has certain characteristics, describing its future payoff: Each asset i has an expected rate of return μ_i per monetary unit, which is paid at time $t_1 \in$ R, $t_1 > t_0$. Let $\mu = [\mu_1, \mu_2, \ldots, \mu_n]^T$. This means if we take a position in $y \in$ R units of asset 1 at time t_0 our expected payoff in t_1 will be $\mu_1 y$ units. Let σ_i be the standard deviation of the return of asset Si. For $i \neq j$, ρ_{ij} denotes the correlation coefficient of the returns of asset S_i and S_j. The correlation coefficient $\rho_{ii} = 1$. Let $\zeta = (\sigma_{ij})$ be $n \times n$ symmetric covariance matrix with $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ for $i \neq j$, and $i, j \in \{1, ..., n\}$. In this notation σ_{ii} is the variance of asset i-th's rate of return and σ_{ij} is the covariance between asset i-th's rate of return and asset j-th's rate of return.

The binary integer programming problem entails the task of minimizing a quadratic objective function while still satisfying linear constraints in the form of equalities and inequalities. In the optimal solution, each variable can be assigned a binary value of either 0 or 1. In scenarios involving multi-criteria optimization, many criteria are simultaneously considered, and it is usually impossible for a single solution to meet all the criteria requirements. It is essential to find a compromise solution that satisfies the decision-preference makers.

A portfolio is defined by a vector $x := (x_1, ..., x_n) \in R^n$, which contains the proportions $x_i \in R$ of the total funds invested in crypto currencies $i \in \{1, ..., n\}$.

We developed the crypto mean-variance optimization model as follows:

$$\textbf{min}\ \text{Xcrypto} = 2^{-1}\text{Xcrypto}\ ^{T}\alpha\ \text{Xcrypto} \tag{1}$$

s. c.

$$\lambda^{T} X_{crypto} \geq \exp R \tag{2}$$

$$lb \leq X_{crypto} \geq ub \tag{3}$$

$$X_{crypto} \geq 0 \tag{4}$$

$$\sum_{i=1}^{n} (x_i) = 1 \tag{5}$$

where $\alpha^{T} \in \mathbb{R}^{m \times n}$, $b = \mathbb{R}^m$, $\lambda \in \mathbb{R}^{n \times n}$ are given, and $x \in \mathbb{R}^n$. Crypto quadratic programming models are a type of nonlinear optimization problem, with some forms being specific instances of linear programming problems.

Quadratic programming components are frequently observed in optimization models. Recall that x is a convex function, which is the objective function (1). R Recall that, when ξ is a positive semi-definite matrix, i.e.

when $\boldsymbol{x}^{t}\,\lambda_{y} \geq \mathbf{0}$ for all *x.* The feasible set is convex because it is a polyhedral set (defined by linear constraint). Consequently, when λ is positive quasi-definite and is positive semi-definite, the Quadratic problem (1) is a convex optimization task. Thus, its globally optimum solutions also happen to be its local optimal ones.

3 The experimental model of cryptocurrencies

The simulation framework is constructed utilizing the aforementioned mathematical methodology, comprising a total of fourteen cryptocurrencies for the year 2023. The geometric mean, correlation matrix, and covariance matrix were computed using actual historical return data for these cryptocurrencies. Cryptocurrencies were grouped into three sets, based on criteria with similar returns. The proposed approach is tested for its applicability and effectiveness using the actual daily stock closing price time series of cryptocurrencies from January 1, 2023, to December 30, 2023, as referenced in [26] and [27]. The article [28] presents sustainable portfolios that provide solutions to enhance decision-making. These portfolios are formulated as quadratic programming (QP) problems, where the objective is to optimize the portfolio. The optimization is subject to the constraint that the sum of returns for the overall portfolio must equal a specified amount [29].

The covariance matrix was calculated and the Coefficients from it were used for the formulation of the problem. The crypto portfolio is data set as a QP:

$$\lambda^{T}\boldsymbol{X} = 0.347154\boldsymbol{X_1} + 2.376808\boldsymbol{X_2} + 0.261502\boldsymbol{X_3} + 4.025126\,\boldsymbol{X_4} + 2.053903\boldsymbol{X_5} + 0.539857\boldsymbol{X_6} + 0.759164\,\boldsymbol{X_7} + 3.856372\boldsymbol{X_8} + 2.683245\boldsymbol{X_9} + 5.841272\boldsymbol{X_{10}} + 7.167325\,\boldsymbol{X_{11}} + 4.053903\boldsymbol{X_{12}} + 3.539857\boldsymbol{X_{13}} + 0.759164\,\boldsymbol{X_{14}} + 1.783269\,\boldsymbol{X_{14}} \geq \mathbf{EXP}\ R$$

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} = 1$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{14} \geq \mathbf{0}$$

The quadratic problem was solved using Matlab's solver. The Firefly method was adapted to address the portfolio problem outlined in references [30-32]. The arrangement method was utilized to guarantee that the total weight of all assets in the portfolio is identical to one, instead of perceiving it as a limitation [33-35]. The problem is resolved thrice using different anticipated rates of return expressed as fractions. The return rates are 0.0050%, 0.0060%, and 0.0070%.

An enhanced firefly algorithm (EFA) for crypto selection

The formulated model (1) - (5) solves a difficult NP-hard issue of nonlinear programming. Conventional robust optimization strategies may not be able to achieve the optimal solution. A novel approach is proposed to efficiently solve the portfolio model by utilizing an improved algorithm that combines the Firefly algorithm and Tabu search for portfolio selection. This approach combines the capability to discover a globally optimal solution (in the case of a multimodal objectives function) with the accurate determination of the optimal solution by reducing the size of the mesh to a predefined tolerance via Tabu search.

The enhanced firefly algorithm for crypto portfolio selection is presented as follows:

Step 1. Determine the EFA parameters: α, β_0 and γ. Set iteration limit – *itlim*. Set diversification limit – *divlim*.

Set iteration counter k=0 and set diversification counter

divcount=0.

Step 2. Initialize fireflies' positions $\{P^k(1),..., P^k(S)\}$, using the three-stage initialization strategy

While (there is improvement of at least one firefly brightness repeat):

Step 3. For each firefly $P^k(i)$ find the brightest firefly it can see.

Step 4. Calculate the new fireflies' positions and update the fireflies' swarm. Update iteration counter: $k = k+1$. Check the stopping criteria and if it is met - go to Step 6.

End While

Step 5. If mod (k/100) = 0, start the *tabu search* procedure.

Step 6. Show the best obtained solution to the decision maker.

Step 7. Check the stopping criteria. If any of the stopping criteria is met - go to Step 8. Otherwise set a diversification search. Update the diversification counter:

divcont = *divcount* + 1.

Step 8. END.

A diversification strategy is especially applicable in instances when the optimal solutions can only be achieved by overcoming specific obstacles that require making actions with lower evaluations. To determine suitable strategies for overcoming obstacles, a memory function can be developed to categorize the relative desirability of different actions within a specific range.

The concept of "move distance" arises from the observation that certain moves result in more significant alterations to the existing solution compared to others. Within the realm of integer programming, the extent to which a certain action affects the relative feasibility or infeasibility of specific constraints, or modifies the value of certain dependent variables, can serve as the foundation for establishing a measure of distance.

4 An EFA solved Crypto portfolio

In this part, portfolio optimization is performed using multi-objective meta-heuristic algorithms (firefly algorithms and tabu search). Therefore, in this approach, there is no limit on the objective function X and both forms are considered minimum. The parameter settings for EFA are as follows: x = 20 (population size), γ = 2, β_0 = 2, α = 0.2, CR=0.2 and F ∈ [0.2 0.8]. The EFA was run with 20 iterations and 20 populations, and when looking at Figure 1, minimum risk results, EFA found the lowest risk respectively 0.000624538.

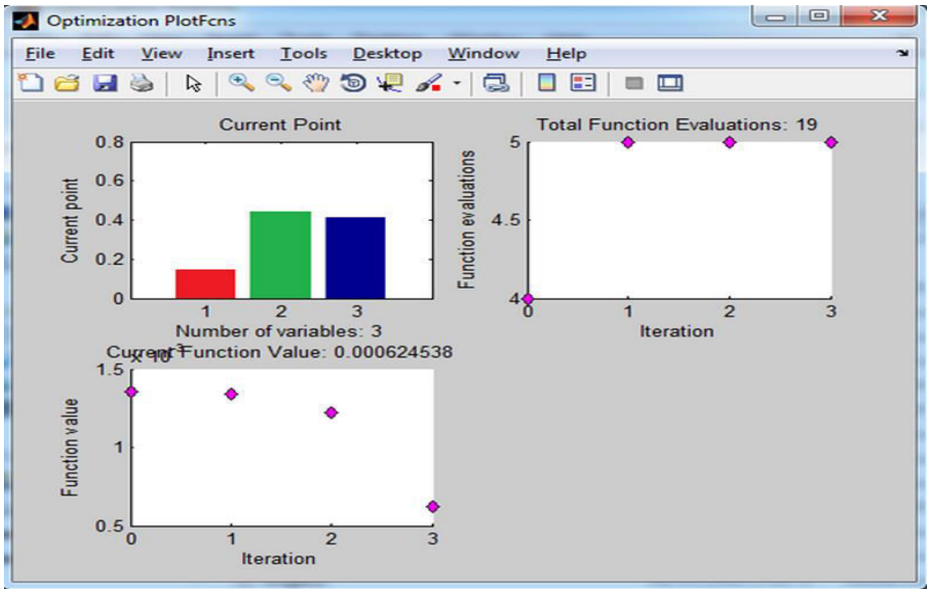


Fig. 1. Optimal crypto portfolio with 0.0050 % expected return

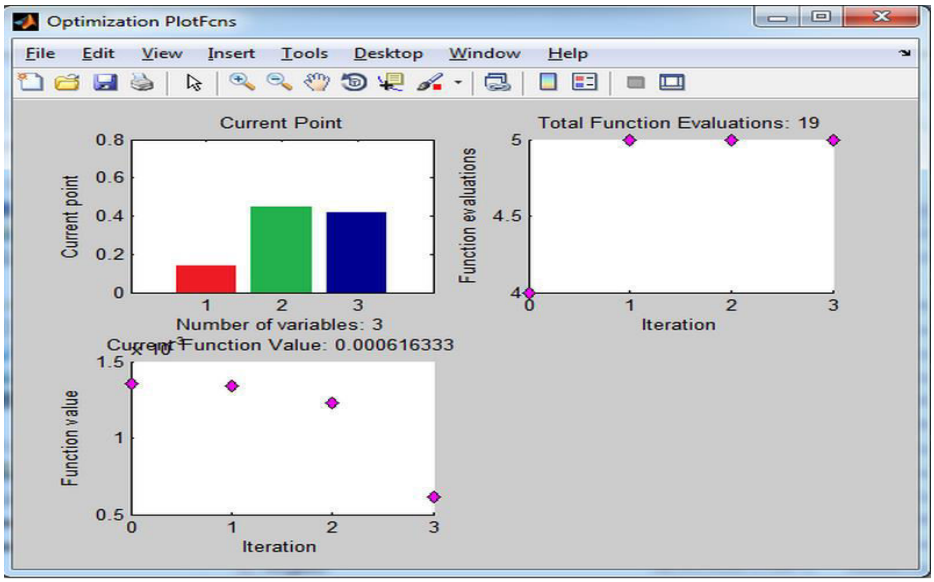


Fig. 2. Optimal crypto portfolio with 0.0060 % expected return

The portfolio strategies for risk-minimizing investors for nineteen objective function value calculations and three iterations for the enhanced firefly algorithm are represented in Figure 2. When we searched the expected return of portfolio with 0.0060 %, the EFA found the lowest risk 0.000616333.

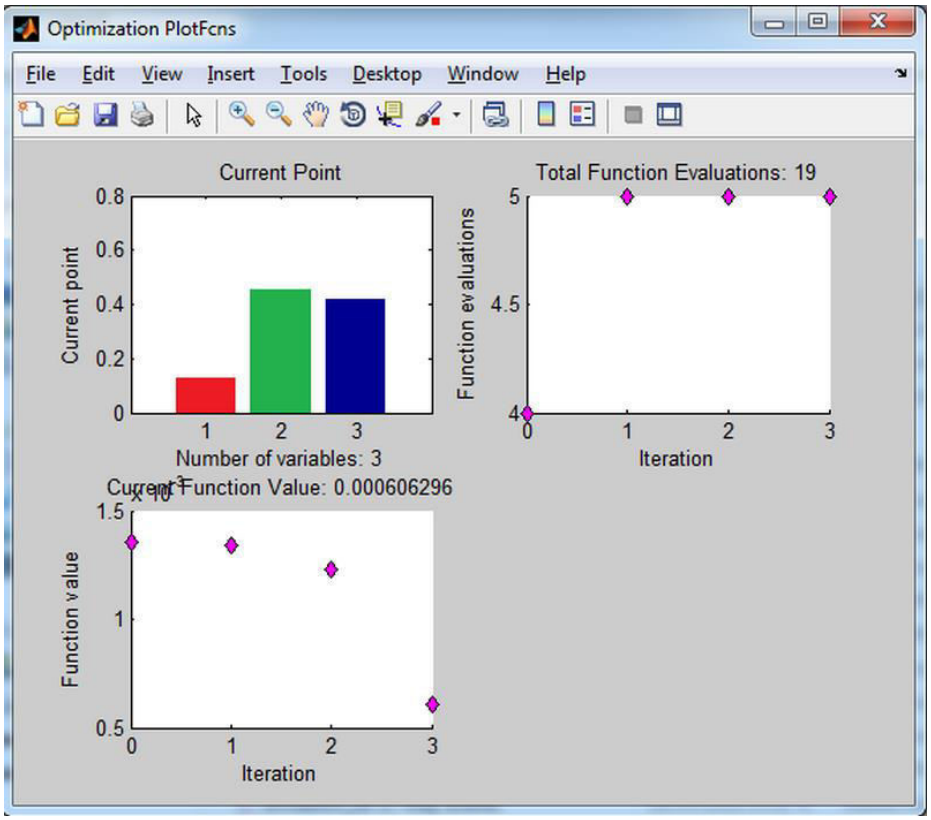


Fig. 3. Optimal crypto portfolio with 0.0070 % expected return

Table 1. Crypto currency optimal portfolio values.

Objective function	**Expected return [%]**	**Total objective function evaluations**	**Optimal Portfolio**		
			Set 1 [%]	Set 2 [%]	Set 3 [%]
6.624573826645712E-4	0.0050	19	5,7319	49,0356	45,2325
6.616333335240 45E-4	0.0060	19	5,1549	49,4163	45,4288
6.606296231159544E-4	0.0070	19	4,5436	49,8173	45,6391

The multi-objective meta-heuristic techniques yielded the following findings for risk minimization: the FA approach achieved the lowest risk value of 0.00604193 with a

return of 0.00185234, while the TS method had a higher risk value of 0.007385421 but a higher return of 0.00328731. It exhibited greater efficiency compared to alternative approaches. However, while aiming to maximize the expected return value, the EFA strategy offers the highest return value of 0.0070 with a comparatively lower risk value of 0.000624538. Our result suggests that EFA has better performance in optimization through cryptocurrency portfolio because it gives a higher return in size based on 0.0070, with FA showing 0.00185234 and TS 0.00328731. When minimizing the risk of the portfolio, the indicators are FA 0.00604193, TS 0.007385421, and EFA 0.000624538. This indicate that employing intelligent technologies can provide financial investors with less risk and increased rewards. Hence, this article concludes that multi-objective meta-heuristic algorithms can assist financial investors in selecting the appropriate portfolio by analyzing the outcomes.

5 Conclusion

Block chain technology and crypto currencies have already had a significant impact on supply chain management and financial sectors. It is projected that this trend will continue in 2024. Advancements in blockchain technology may enable the creation of more efficient and secure systems for trading digital assets, managing digital identities, and implementing decentralized financing (DeFi). The research presented an improved methodology that combines two meta-heuristic algorithms, FA and TS, for globally optimizing green investments in cryptocurrency portfolios while considering restrictions. The updated technique for solving the crypto portfolio achieves the optimal balance between return and diversification.

The EFA was evaluated using a dataset from 2023 that included fourteen actual cryptocurrencies and historical data. The EFA, which is experiencing tremendous growth and demonstrating great effectiveness, indicates the valuable potential of this method and its fundamental concepts. The implications of this have the potential to create more profound links between artificial intelligence and mathematical optimization, which present promising opportunities for further research. The approach is convenient for conducting early investigations because to its ability to easily launch rudimentary implementations with minimal effort and the option to expand on them as necessary. As more improvements are made, the use of learning methods like green analysis allows for a more comprehensive utilization of the two main opposing forces - represented by the interaction between limitations and desired criteria, and between strategies of intensification and diversification. These endeavours are likely to result in more efficient modifications and to create opportunities for further investigation and application.

Acknowledgments. This work is supported by the Bulgarian National Science Fund by the project "Mathematical models, methods, and algorithms for solving hard optimization problems to achieve high security in communications and better economic sustainability", KP-06-N52/7/19-11-2021.

References

1. A. Leng et al., Blockchain-empowered sustainable manufacturing and product lifecycle management in industry 4.0: A survey, Renewable and Sustainable Energy Reviews, Volume 132, 2020, https://doi.org/10.1016/j.rser.2020.110112hu-Cherng Fang and Sarat Puthenpura. 1993. Linear optimization and extensions: theory and algorithms. Prentice-Hall, Inc., USA.
2. M. L. Di Silvestre et al., Blockchain for power systems: Current trends and future applications, Renewable and Sustainable Energy Reviews, Volume 119, 2020, https://doi.org/10.1016/j.rser.2019.109585.
3. D.G. Baur et al., Bitcoin: Medium of exchange or speculative assets, Journal of International Financial Markets, Institutions and Money, Volume 54, 2018, Pages 177-189, https://doi.org/10.1016/j.intfin.2017.12.004.
4. J.W. Goodell et al., Diversifying equity with cryptocurrencies during COVID-19, International Review of Financial Analysis, Volume 76, 2021, https://doi.org/10.1016/j.irfa.2021.101781.
5. E. Bouri et al., On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier, Finance Research Letters, Volume 20, 2017, Pages 192-198, https://doi.org/10.1016/j.frl.2016.09.025.
6. C.L. Bastian-Pinto et al., Hedging renewable energy investments with Bitcoin mining, Renewable and Sustainable Energy Reviews,Volume 138, 2021, https://doi.org/10.1016/j.rser.2020.110520.
7. D.G. Baur et al., A crypto safe haven against Bitcoin, Finance Research Letters, Volume 38, 2021, https://doi.org/10.1016/j.frl.2020.101431.
8. X. Li et al., The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin, Decision Support Systems, Volume 95, 2017, Pages 49-60, https://doi.org/10.1016/j.dss.2016.12.001.
9. Markowitz, H.M., Portfolio selection", Journal of Finance, 1952, 7(1), P. 77-91. Doi: 10.2307/2975974
10. Schaerf, A., Local search techniques for constrained portfolio selection problems, Computational Econom-ics, 2002, 20(3), P. 177-190. Doi: 10.1023/A:1020920706534.
11. Yudong Zhang, Shuihua Wang, Genlin Ji, "A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications", Mathematical Problems in Engineering, vol. 2015, Article ID 931256, 38 pages, 2015. https://doi.org/10.1155/2015/931256
12. Yang, X. S., Nature-Inspired Metaheuristic Algorithms, Luniver Press, (2008).
13. Yang,X. S., Firefly algorithms for multimodal optimization, In: Stochastic Algorithms: Foundations and Applications, SAGA 2009, Lecture Notes in Computer Science, 5792, 2009, pp. 169-178.
14. Yang, X. S., Firefly algorithm, Levy flights and global optimization, In: Research and Development in Intelligent Systems XXVI, (Eds M. Bramer et al.), Springer, London, 2010, pp. 209-218.

15. Glover, F., and C. McMillan, “The General Employee Scheduling Problem: An Integration of Management Science and Artificial Intelligence,” Computers and Operations Research, Vol. 13, No. 5, 563-593, 1986.
16. Jaumard, B., P. Hansen and M. Poggi de Aragao, “Column Generation Methods for Probabilistic Logic,” GERAD, G-89-40, McGill University, November 1989
17. Ryan, J., ed. Final Report of Mathematics Clinic: Heuristics for Combinatorial Optimization, June 1989
18. Skorin-Kapov, J., “Tabu Search Applied to the Quadratic Assignment Problem,” Research Report, HAR-89-001, W. A. Harriman School for Management and Policy, SUNY at Stony Brook, N.Y, May 1989, to appear in ORSA Journal an Computing.
19. Hansen, P. and B. Jaumard, “Algorithms for the Maximum Satisfiability Problem,” RUTCOR Research Report RR ## 43-87, Rutgers, New Brunswick, NJ, 1987.
20. Herz, A., and D. de Werra, “Using Tabu Search Techniques for Graph Coloring,” Computing, Vol. 29, pp. 345-351, 1987.
21. Knox, J. and F. Glover, “Comparative Testing of Traveling Salesman Heuristics Derived from Tabu Search, Genetic Algorithms and Simulated Annealing.” Center for Applied Artificial Intelligence, University of Colorado, July 1989.
22. Malek, M., M. Guruswamy, H. Owens and M. Pandya, “Serial and Parallel Search Techniques for the Traveling Salesman Problem,” Annals of QR: Linkages with Artificial Intelligence, 1989.
23. Benotsmane, R.; Dudás, L.; Kovács, G. Newly Elaborated Hybrid Algorithm for Optimization of Robot Arm’s Trajectory in Order to Increase Efficiency and Provide Sustainability in Production. Sustainability 2021, 13, 8193. https://doi.org/10.3390/su13158193
24. Borissova, D., Dimitrova, Z., Keremedchieva, N. (2023). Software Application to Assist the Publishing Sector: A Tool in MS Excel Environment. In: Rocha, Á., Ferrás, C., Ibarra, W. (eds) Information Technology and Systems. ICITS 2023. Lecture Notes in Networks and Systems, vol 692. Springer, Cham.
25. Guliashki, Vassil, Kirilov, Leoneed and Nuzi, Alsa. "Optimization Models and Strategy Approaches Dealing with Economic Crises, Natural Disasters, and Pandemics – An Overview" Cybernetics and Information Technologies, vol.23, no.4, 2023, pp.3-25. https://doi.org/10.2478/cait-2023-0033
26. Abolmakarem, Shaghayegh and Abdi, Farshid and Khalili-Damghani, Kaveh and Didehkhani, Hosein, A Multi-Stage Machine Learning Approach for Stock Price Prediction: Engineered and Derivative Indices. Available at SSRN: https://ssrn.com/abstract=4074883 or http://dx.doi.org/10.2139/ssrn.4074883
27. Abolmakarem, S., Abdi, F., Khalili-Damghani, K. and Didehkhani, H. (2023), "Futuristic portfolio optimization problem: wavelet based long short-term memory", Journal of Modelling in Management, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JM2-09-2022-0232
28. Stoyanova K. and Balabanov T., "A combination of Broyden-Fletcher-Goldfarb-Shanno (BFGS) and bisection method for solving portfolio optimization problems," 2022 International Conference on Engineering and Emerging Technologies

(ICEET), Kuala Lumpur, Malaysia, 2022, pp. 1-3, doi: 10.1109/ICEET56468.2022.10007369.
29. Stoyanova K. and T. Balabanov, "Optimal Selection of Pharma Stock Portfolios using DEPSO," 2023 24th International Carpathian Control Conference (ICCC), Miskolc-Szilvásvárad, Hungary, 2023, pp. 419-422, doi: 10.1109/ICCC57093.2023.10178900.
30. Tuba M. and Bacanin N., "Upgraded Firefly Algorithm for Portfolio Optimization Problem," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 113-118, doi: 10.1109/UKSim.2014.25.
31. Bacanin, N., & Tuba, M. (2014). Firefly algorithm for cardinality constrained mean-variance portfolio optimization problem with entropy diversity constraint. The Scientific World Journal, 2014. https://doi.org/10.1155/2014/721521
32. Ramshe, M., Gharakhani, M., Feyz, A., & Sadjadi, S. J. (2021). A Firefly Algorithm for Portfolio Optimization Problem with Cardinality Constraint. International Journal of Industrial Engineering and Management Science, 8(1), 24-33. doi: 10.22034/ijiems.2021.297820.1044
33. Lazulfa, Indana. "A Firefly Algorithm for Portfolio Optimization." Journal of the Indonesian Mathematical Society, vol. 25, no. 3, Nov. 2019, pp. 282-291, doi:10.22342/jims.25.3.821.282-291.
34. Shahid, M., Ashraf, Z., Shamim, M. and Ansari, M.S. (2022), "Solving constrained portfolio optimization model using stochastic fractal search approach", International Journal of Intelligent Computing and Cybernetics, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJICC-03-2022-0086
35. Ron, D., “Development of Fast Numerical Solvers for Problems in Optimization and Statistical Mechanics.” Ph.D. thesis, Dept. of Applied Mathematics, The Wiezmann Institute of Science, Rehovot, Israel, 1988.

Comprehensive Approach to Enhancing Digital Accessibility with EaseAccess24

Filip Najdovski[1,2], Patrick Tairi[1], and Biljana Risteska Stojkoska[3]

[1]T-Meeting, Amiralsgatan 20, 211 55 Malmö, Sweden

[2]Saba High School, Anton Popov 1-2, Skopje 1000, North Macedonia

[3]Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Rugjer Boskovic 16, Skopje, 1000, North Macedonia

Abstract. Web accessibility is an essential aspect of the present internet society in which we reside. For a considerable duration, the conduction of research in the field of accessible web technologies has experienced an exponential growth, initiating and amplifying guidelines that shape the accessibility web environment such as WCAG, ADA and ARIA. While this has made a significantly beneficial impact in terms of the development process of these technologies, the establishment of confluence between the stakeholders, which holds vital importance for making these features obtainable for routine patrons in the world of web, had not been achieved. Leading up to 2019, when the European Accessibility Act was adopted to enhance website and mobile app accessibility, with full implementation required by 2025. In this paper, a comprehensive overview of the current implementation of accessible web practices is outlined. Considering the current WCAG Guidelines a suitable solution "EaseAccess24" is being provided. A detailed analysis of its components and impact on the society is indicated, including the features, technologies, functionalities and characteristics to conquer the present challenges.

Keywords: Web Accessibility, WCAG, ADA, ARIA, European Accessibility Act, Accessible Technologies

1 Introduction

Over the recent period, technological advancement in the world of the web has experienced substantial growth, spreading wide across all fields and becoming impossible for one to live without the ability to be part of it. This has led to the emergence of heightened challenges for individuals with certain disabilities, encompassing those with motor impairments, color blindness, dyslexia, visual impairments, ADHD, cognitive impairments, and learning disabilities, in accessing the web [1].

Thus, web accessibility became an indispensable component that a user-friendly web should embed, and with it, a significant array of regulations were established for developers, including but not limited to WCAG [2], ARIA [3], and ADA [4]. This implies that the potential of developing these technologies is exceptional, especially now, having a versatile development stack to build them on the web.

Although there are various technical stacks that can be used to integrate this adaptive software, commonly encountered as optimal practices are React.js, Vue.js, Angular, Node.js, and Express.js. The implementation of the relevant regulations, alongside these technologies in the development of a widget, can facilitate a broad spectrum of applications across an extensive range of fields and within existing web-based software, delivering outstanding benefits.

For instance, implementation in e-commerce will enable easier and more effective navigation and purchasing; in education, the platforms could accommodate diverse learning needs; healthcare platforms could have enhanced usability for patients with accessibility requirements. In media and entertainment, sites will be inclusive for all viewers and when it comes to government websites, it is a must that they comply with accessibility standards to provide equal access to their services, and their public service applications ought to allow access to essential information for everyone.

However, integrating this technology cross-platform so that it functions seamlessly on websites with varying technical structures and content management systems presents a substantial challenge, which requires a lot of problem-solving and creativity. Other challenges involve resource allocation, established managerial practices, and time limitations. Web developers and designers may also lack the necessary knowledge to implement techniques that support accessibility on the web.

The aim of this paper is *(i)* to provide a summary of the existing web accessibility technologies and regulations, *(ii)* to supply a detailed analysis of the effects of these technologies on society and the business community worldwide, and *(iii)* to present our "EaseAccess24 Widget" which serves as a complete universal solution that contributes to society while providing a resolution for the underlying accessibility challenges.

The remainder of this paper is organized as follows: Section II discusses current trends in web accessibility. Section III provides an overview of the technology used in this field. Section IV focuses on our EaseAccess24's accessibility widget. Section V analyzes the social impact and business implications of these developments. Finally, Section VI concludes the paper and proposes future improvements in accessibility.

2 Current Accessibility Trends

Throughout the past years, web accessibility standards have become increasingly sophisticated and advanced, with Web Content Accessibility Guidelines (WCAG) being the most recognized and authoritative. WCAG offers a distinct set of rules, categorized into four foundations: Perceivable, Operable, Understandable and Robust (POUR), with the goal of improving web accessibility for people with disabilities. Every one of these bases has a unique specification to handle the range of accessibility issues [5].

2.1 The Perceivable Principle

To exemplify, under the base of the perceivable principle, guidelines focus on enabling content to the senses of sight and hearing. For a website to be considered perceivable, the content that's uploaded on the website must reach the users through the senses they rely on when receiving and interpreting information.

Guideline 1.1 (Text Alternatives) implies providing text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols, or simpler language.

Guideline 1.2 (Time-based Media) The purpose of this guideline is to provide access to time-based and synchronized media, including media that is: audio-only, video-only, audio-video, audio and/or video combined with interaction.

Guideline 1.3 (Adaptable) indicates creating content that can be presented in different ways, for example, simpler layouts without losing information or structure, while accommodating different needs.

Guideline 1.4 (Distinguishable) relates to making sure the base content is easy to discern from backgrounds and other decorations. The most commonly used example is color, including but not limited to color contrast and use of color to convey instructions.

2.2 The Operable Principle

The operable principle ensures that interface components and navigation are usable for individuals encountering difficulties.

With guideline 2.1 (Keyboard Accessible) conveying that all functionality can be achieved using the keyboard. It can be accomplished by keyboard users, by speech input which then creates keyboard input, by mouse using on-screen keyboards, and by a wide variety of assistive technologies that create simulated keystrokes as their output.

Guideline 2.2 (Enough Time), covering situations in which functionality may have a time limit, a given example could be online purchases which sometimes need to be completed within a time limit for security reasons.

Guideline 2.3 (Seizures), pertains to content that, if left unaltered, has the potential to trigger seizures in individuals with conditions like epilepsy or induce physical reactions, such as dizziness, in those with vestibular disorders.

Guideline 2.4 (Navigable), the intent of this guideline is to help users find the content they need and allow them to keep track of their location. And under Guideline 2.5 (Input Modalities), ensuring that users are able to interact with digital technology using different input methods beyond a keyboard or mouse such as touchscreen, voice, device motion, or alternative input devices.

2.3 The Understandable Principle

The understandable principle states that information and the operation of the user interface must be understandable, supporting different criteria.

Guideline 3.1 (Readable), focusing on making text content easily approachable and readable.

Guideline 3.2 (Predictable), targeting the intuitive level of interfaces.

Guideline 3.3 (Input Assistance), centering around supporting users to enter correct information or information in the correct way with the minimum possible mistakes.

2.4 The Robust Principle

The robust principle states that content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies.

Guideline 4.1 (Compatible), focusing on making content as compatible as possible for current and future user agents.

3 Technology Overview

The technology environment for this field is rapidly evolving. Among the essential technologies are screen readers. For instance, JAWS (Job Access With Speech) and NVDA (NonVisual Desktop Access) are tools that translate text and other visual data into speech or even braille, which enables visually impaired users to access the digital world [6] [7].

3.1 Assistive Technologies

Tera is another extraordinary technology, a text-to-speech and a speech-to-text app designed to assist individuals with special disabilities. It enables users to seamlessly convert audio to text and text to audio during phone calls, enhancing communication with AI [8] [9]. Another example of technology that's utilized in this industry is Dragon NaturallySpeaking, voice recognition software by Nuance Communications. This software allows individuals with certain limitations to mobility to use computers and other devices by spoken commands. Using these technologies, people may complete complicated actions and operate the devices by opening apps, creating or editing emails and documents, controlling the mouse, and other complex tasks while remaining hands-free [10]. UserWay and accessiBe are two prominent platforms that offer AI-powered accessibility solutions for websites, helping organizations meet compliance standards like WCAG and ADA. Both services provide automated tools, such as screen reader optimization, keyboard navigation, and customizable accessibility settings, to enhance user experience for individuals with disabilities [11] [12].

3.2 Evaluation Tools for Web Accessibility

Furthermore, there are evaluation tools such as WAVE (Web Accessibility Evaluation Tool) and AXE by Deque Systems, that can constantly assist developers in identifying and addressing issues. Including automated testing capabilities, they are contributing to the insurance of web accessibility standards across the web [13].

To enhance web accessibility even further, developers utilize Accessible Rich Internet Applications (ARIA), which is a set of attributes that provide additional context and functionality to HTML elements, improving the interaction between the user and the web.

3.3 The Role of AI in Web Accessibility

Additionally, AI holds special potential to make the web more accessible. Tools powered by artificial intelligence can automate the alternative (alt) text for images, transcribe audio content, and predict accessibility issues before they impact users. For instance, Microsoft's AI capabilities can automatically generate descriptive alt text for images, improving the accessibility of visual content for screen reader users.

Another interesting use of AI is analyzing user behavior to personalize accessibility features with machine learning algorithms, making web interactions more intuitive and adaptive to the needs of specific individuals. To exemplify, Google AI leverages machine learning to provide real-time language translation and adaptive content, enhancing the web experience for users with diverse needs. AI can also assist in developing more robust accessibility evaluation tools.

4 EaseAccess24 - Accessibility Widget

EaseAccess24 is an innovative widget that aims to significantly enhance the user experience for individuals with disabilities when navigating the web. By embedding a comprehensive suite of accessibility tools and features, the widget addresses the diverse needs of users who encounter challenges when accessing digital content on the internet. The widget complies rigorously with guidelines such as the Web Content Accessibility Guidelines (WCAG) and the Americans with Disabilities Act (ADA), guaranteeing that the required level of accessibility is achieved. The commitment to comply with these guidelines is not limited to merely facilitating an inclusive web environment but also extends to supporting website owners in adhering to legal standards, hence scaling their outreach to a wider audience. Adaptivity and integration are crucial for achieving accessibility. EaseAccess24 integrates seamlessly with websites, offering minimal configuration. Fig. 1 provides a preview of the interface of the widget we developed in its beta version.

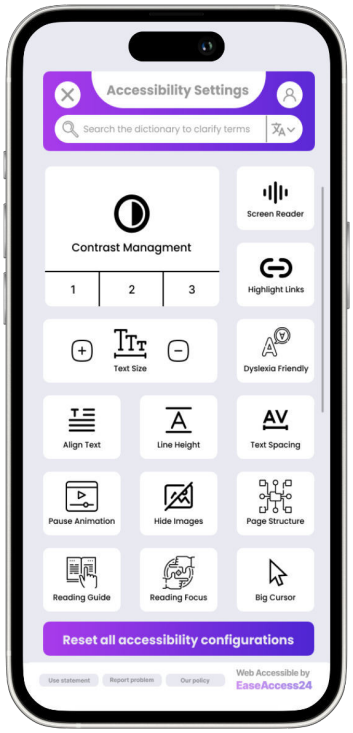


Fig. 1. Preview of the EaseAccess24 widget in its beta version.

4.1 Customizable Accessibility Features

The widget provides a variety of customizable options to achieve accessibility, including color contrast adjustment, text sizing and other text adjustments, alternative text descriptions, keyboard control, screen reader, cursor options, and page navigation support. These features are combined in an intuitive interface, allowing users to modify their viewing preferences, ensuring that web content is accessible and user-friendly. The comparison between our EasyAccess24 and the two current popular platforms on the market is given on Fig. 2. It is evident, our widget provides the same features as the concurrent platforms, but is superior in terms of customization options and auditing. Providing an option to make an individual profile in which most of the functionalities can be customized.

	accessiBe	UserWay	EaseAccess24
AI-Powered Accessibility	AI addresses screen reader adjustments and keyboard navigation, but lacks proactive accessibility insights.	Detects issues and provides basic fixes, mainly focused on contrast and alt text.	Advanced AI for content analysis, alt text, screen reader, and proactive accessibility issue identification.
Compliance Standards	WCAG 2.1, ADA, Section 508, EN 301549.	WCAG 2.1, ADA, Section 508, EN 301549.	WCAG 2.1, ADA, Section 508, EN 301549.
Accessibility Audits	Continuous audits, but mostly relies on AI for automatic corrections without comprehensive real-time insights.	Automated audits, but may require manual intervention for some fixes.	Real-time, comprehensive WCAG 2.1 audits with instant reporting and solutions for continuous compliance improvement.
Customization Options	Basic customization, with predefined accessibility profiles but less flexibility for design modifications.	Limited customization focused on widget color and basic settings.	Full customization of widget design, colors, and user-specific profiles across all websites using the widget.
UI/UX of the Widget	Easy to use but offers limited control over design and lacks deeper personalization for specific user needs.	Simple interface but lacks advanced customization options and user-specific profiles.	User-friendly, fully customizable interface with seamless navigation tailored to individual user needs and preferences.

Fig. 2. Comparison table between accesiBe, UserWay, and EaseAccess24.

4.2 Seamless Integration and Adaptivity

Adaptivity and integration are essential for achieving effective accessibility. EaseAccess24 is designed with these principles in mind, offering a solution that integrates seamlessly with websites and requires minimal configuration. For developers, this means a straightforward implementation process that minimizes setup time and complexity. The ease of integration ensures that accessibility features can be quickly and efficiently incorporated into existing digital environments, allowing developers to enhance website accessibility without extensive

adjustments or disruptions. Fig. 3 provides a preview of the widget's workflow once initially integrated correctly. This web accessibility tool allows users to customize settings, enabling real-time adjustments through secure data processing and AI enhancements, which ensures compliance and continuously optimizes the experience.

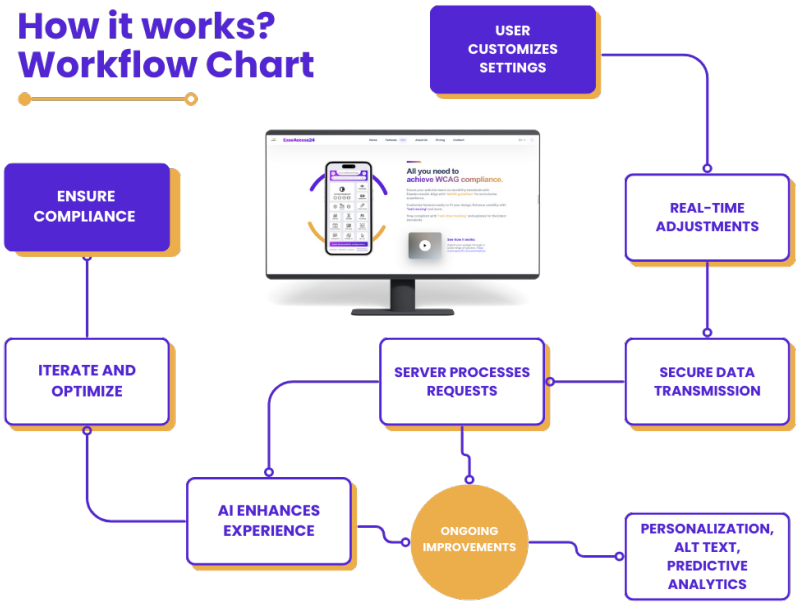


Fig. 3. EaseAcess24 Widget Workflow Chart.

4.3 Secure Data Management and AI-Enhanced Accessibility

User preferences are securely maintained in a protected database with access controls in place to prevent unauthorized access. The server manages these requests and updates the database accordingly. EaseAccess24 integrates artificial intelligence to further improve accessibility features. AI algorithms analyze user behavior and preferences to provide personalized recommendations for accessibility settings. Machine learning models predict optimal text sizes or contrast levels based on user interaction patterns. Additionally, AI is used to automate the creation of descriptive alt text for images, enhancing accessibility for users who rely on screen readers.

5 Social Impact and Business Implications

The implications of web accessibility technology on society are vast, as it directly affects the quality of life for individuals with disabilities. The wider implementation of website accessibility will ensure that society moves closer to digital inclusivity, where everyone, regardless of physical or cognitive abilities, can participate fully in the digital world. This inclusivity is vital as it promotes equal access to information, services, and opportunities online.

5.1 Web Accessibility and Disability Statistics

Global Disability Statistics: Over 1.3 billion people, or about 16% of the world's population, experience some form of disability. In the European Union (EU), approximately 101 million people, or around 27% of the population over 16, live with some form of disability. This includes visual, auditory, cognitive, and motor impairments [14][15].

Internet Usage Among People with Disabilities: According to The United States Department of Labor (2022), about 65% of people with disabilities in the US used the internet regularly, compared to 90% of those without disabilities. This statistic underscores the digital divide and highlights the importance of web accessibility in bridging this gap [16].

Economic Impact of Accessibility: The estimated disposable income of people with disabilities in the EU is approximately €1.3 trillion. This represents a significant consumer market that businesses can access by improving digital accessibility [17].

5.2 Web Accessibility Legal Landscape

Compliance: A survey conducted by the European Commission in 2022 found that only 23% of public sector websites and 12% of private sector websites in the EU fully comply with WCAG 2.1 standards, indicating a significant gap in web accessibility that needs to be addressed. Furthermore, in 2022, there were 3,255 ADA Title III website accessibility lawsuits filed in U.S. federal courts, a 14% increase from the previous year, emphasizing the growing legal risks for businesses that fail to meet accessibility standards [18][19].

Global Accessibility Standards: The Web Content Accessibility Guidelines (WCAG) are recognized as the international standard for web accessibility. The WCAG 2.1 guidelines, which include 78 success criteria, are used by most organizations to ensure compliance [2].

5.3 Business Benefits of Accessibility

Improved Customer Engagement: Websites that implement accessibility features see a 12% increase in overall traffic and a 15% reduction in bounce rates on average, according to case studies from other companies [20].

Accessibility and Brand Loyalty: A study by Forrester Research found that 69 of customers with disabilities are more likely to trust a brand that demonstrates a commitment to digital accessibility [21].

Market Expansion Through Accessibility: Businesses in Europe that have implemented web accessibility features report an average increase of 10-15% in their customer base from the disabled community. Additionally, accessible websites have been shown to increase overall customer satisfaction by 20% [22].

Public Perception and Accessibility: A study by the European Consumer Organisation (BEUC) found that 67% of European consumers are more likely to engage with brands that demonstrate a commitment to accessibility. This reflects the growing importance of accessibility in brand loyalty [23].

5.4 Digital Inclusion in Education

Digital Inclusion and Education: Acording to European University Association, in 2022 in the EU, 10% of university students report having a disability. The adoption of accessible e-learning platforms is crucial to ensuring that these students can fully participate in higher education. However, only 40% of European universities have fully accessible online platforms, indicating a need for improvement [24].

6 Conclusion

The advancement of web accessibility technologies, guided by established standards such as WCAG, ADA, and ARIA, has become a critical component of digital inclusivity. The integration of these technologies across various industries, including education, healthcare, and e-commerce, demonstrates their broad societal impact. Despite the challenges of implementation, such as cross-platform integration and developer expertise, the benefits are clear. The development of tools like EaseAccess24 and the growing role of AI in accessibility highlight the potential for continued progress. As web accessibility becomes increasingly mandated by legal frameworks like the European Accessibility Act, the ongoing commitment to these standards will be essential in ensuring equitable access to digital content for all users.

7 Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje.

References

1. Zhang, A., 2016. *Web Accessibility Challenges*. International Journal of Advanced Computer Science and Applications, 7(1), 234-242. DOI: https://doi.org/10.14569/IJACSA.2016.071023
2. World Wide Web Consortium (W3C), 2024. *Web Content Accessibility Guidelines (WCAG) 2.1*. Retrieved from https://www.w3.org/WAI/WCAG21/
3. World Wide Web Consortium (W3C), 2023. *Accessible Rich Internet Applications (ARIA)*. Retrieved from https://www.w3.org/TR/wai-aria/
4. U.S. Department of Justice, 2023. *Americans with Disabilities Act (ADA)*. Retrieved from https://www.ada.gov/
5. World Wide Web Consortium (W3C), 2018. *Web Content Accessibility Guidelines (WCAG) 2.1*. Retrieved from https://www.w3.org/WAI/WCAG21/
6. Freedom Scientific, 2022. *JAWS (Job Access With Speech)*. Retrieved from https://www.freedomscientific.com/products/software/jaws/
7. NV Access, 2022. *NonVisual Desktop Access (NVDA)*. Retrieved from https://www.nvaccess.org/download/
8. T-Meeting, 2017. *Tera: Text-to-Speech and Speech-to-Text App*. Retrieved from https://www.tmeeting.com/our-products/tera
9. T-Meeting, 2022. *Tera - The World's Most Advanced Speech-to-Text App for Telephone Calls*. Retrieved from https://mb.cision.com/Public/21671/3668894/be232ca2386832de.pdf
10. Nuance Communications, 2021. *Dragon NaturallySpeaking: Voice Recognition Software*. Retrieved from https://www.nuance.com/dragon.html
11. UserWay, 2024. *Dragon NaturallySpeaking: Voice Recognition Software*. Retrieved from https://userway.org/content/
12. accessiBe, 2024. *Dragon NaturallySpeaking: Voice Recognition Software*. Retrieved from https://https://accessibe.com/
13. Deque Systems, 2024. *axe Accessibility Scanner*. Retrieved from https://www.deque.com/axe/
14. World Health Organization, 2021. *Disability and Health*. Retrieved from https://www.who.int/health-topics/disability
15. European Disability Forum, 2022. *Disability Rights in the EU*. Retrieved from https://www.edf-feph.org/
16. U.S. Department of Labor, 2022. *The Disability Digital Divide: An Overview*. Retrieved from https://www.dol.gov/sites/dolgov/files/ODEP/pdf/disability-digital-divide-brief.pdf
17. European Commission, 2020. *The Economic Impact of Digital Accessibility*. Retrieved from https://ec.europa.eu/
18. European Commission, 2022. *Web Accessibility Compliance in the EU*. Retrieved from https://ec.europa.eu/
19. Seyfarth Shaw LLP, 2023. *ADA Title III Lawsuits by the Numbers*. Retrieved from https://www.adatitleiii.com/2024/06/federal-court-website-accessibility-lawsuit-filings-took-a-dip-in-2023/

20. UserWay, 2023. *Case Studies in Web Accessibility*. Retrieved from https://userway.org/
21. Forrester Research, 2021. *The Impact of Accessibility on Brand Loyalty*. Retrieved from https://www.forrester.com/
22. AbilityNet, 2023. *Accessibility in Business: A European Perspective*. Retrieved from https://www.abilitynet.org.uk/
23. European Consumer Organisation (BEUC), 2023. *Consumer Attitudes Towards Accessibility*. Retrieved from https://www.beuc.eu/
24. European University Association, 2022. *Accessibility in European Higher Education*. Retrieved from https://eua.eu/

Session 2, 3

The Impact of Packet Loss Recovery Mechanisms and Network Performance in SD-WAN Compared to Traditional WAN

Mitko Jankuloski[1] and Stojan Kitanov[2][0000−0002−7222−1078]

[1] London Metropolitan University, Skopje Metropolitan College, Skopje, Skopje, R. N. Macedonia
mitkoj.ohrid@gmail.com
[2] Mother Teresa University, Faculty of Information Sciences, Skopje, R. N. Macedonia
stojan.kitanov@unt.edu.mk

Abstract. The evolution of Wide Area Networks (WANs) has brought about significant changes in how organizations manage network performance and Quality of Service (QoS). Nowadays, numerous applications run over these networks, each demanding specific QoS criteria such as packet loss, latency, and jitter. Some of these new applications require exceptionally high performance, making it difficult to manage traffic accurately and meet Service Level Agreements (SLAs). Traditional WANs are finding it increasingly difficult to meet these evolving SLA demands. In this context, SD-WAN presents a promising solution to address these challenges. One of the key areas where SD-WAN shows significant improvement over traditional WANs is in packet loss recovery mechanisms. The experimental results of the analysis using the feature Forward Error Correction (FEC) as a packet loss recovery mechanism presented in this paper demonstrate that Cisco SD-WAN significantly improves overall network performance and QoS compared to traditional WAN systems. The network emulation software EVE-NG is used as a primary tool for this research experiment.

Keywords: EVE-NG · Forward Error Correction (FEC) · Packet loss · SD-WAN.

1 Introduction

Network technologies have had a huge leap in technological development in recent years. With the rapid expansion of networks and the emergence of increasingly sophisticated applications, many enterprises and organizations need reliable and high-performing WANs to effectively transmit critical data between their branches, data centers, SaaS, and other cloud-based applications.

Therefore, Software Defined Wide Area Network (SD-WAN) is quickly becoming an attractive solution for enterprise networks as it can accommodate these desired capabilities. SD-WAN is a promising technology that has recently received a lot of attention from industry and academia [1].

The objective of this paper is to analyze the impact of SD-WAN compared with traditional WAN regarding Quality of service (QoS) and network performances. The focus will be on the packet loss as a key performance parameter.

This paper analysis is based on an experimental SD-WAN design using Cisco methods for implementation. This experimental SD-WAN design is implemented in a lab environment using the network emulation software EVE-NG.

The subject of the research within this paper is the performance analysis of Cisco SD-WAN. The research aims to show that Cisco SD-WAN concepts can have a positive impact on the performance parameters compared to Traditional WAN.

The paper is organized as follows. After the Introduction provided in section 1. Section 2 provides an overview of packet switching mechanisms in traditional WANs and introduces Cisco Express Forwarding as a new packet switching method. Section 3 delves into Cisco SD-WAN technology, outlining the main components of each plane within this solution. Section 4 offers an overview of packet loss and investigates the Forward Error Correction (FEC) feature in Cisco SD-WAN as a method to prevent packet loss. This section also covers the research methodology, a brief explanation of the tools used, and the experimental SD-WAN design. Additionally, it includes the interpretation of the results and analysis of the experiments. Furthermore, this section describes the retransmission mechanism for packet loss recovery in traditional WANs and presents the mathematical model for the FEC adaptive feature in SD-WAN. In addition latency and jitter are also discussed. Finally, Section 5 concludes the paper with a summary of the findings and suggestions for future work.

2 Overview of Traditional WAN

During the early phases of Cisco router development, the mechanism for packet switching was referred to as process switching. However, as networking technology improved, Cisco developed new mechanisms for packet switching known as fast switching and Cisco Express Forwarding (CEF) in order to further improve packet handling capabilities, and to meet the escalating demands of evolving network infrastructures [2].

CEF provides optimization of network performance and scalability using the forwarding information known as Forwarding Information Base (FIB), and the cached adjacency information known as Adjacency Table. It's very less CPU intensive, provides faster speed and updates its FIB and Adjacency immediately. It demands minimal CPU resources, delivers rapid speeds, and quickly updates its FIB and Adjacency table.

The FIB is formed directly from the routing table and stores the next-hop IP address for every destination within the network. Whenever there is a routing or topology change within the network, the IP routing table is updated, and these changes are updated in the FIB. CEF uses the FIB to make switching decisions based on IP destination prefixes.

The adjacency table stores the directly connected next-hop IP addresses and their corresponding next-hop MAC addresses, along with the MAC address of the egress interface. The adjacency table is sourced with data from the ARP table.

An overview of the CEF process flow is illustrated on Fig. 1.

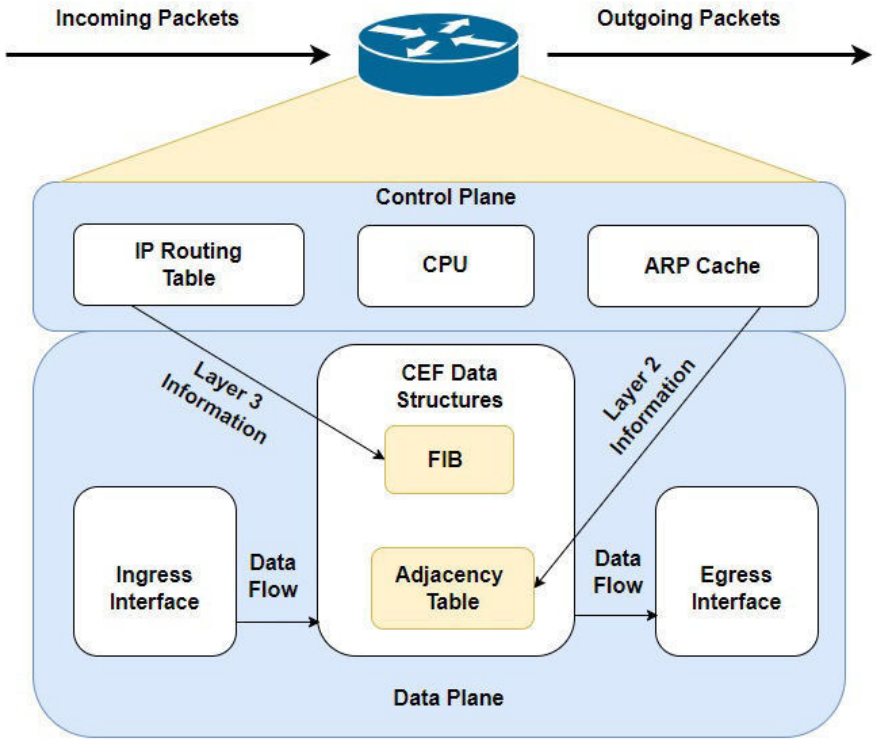


Fig. 1. CEF Process Flow.

When a packet enters the router, packets undergo a transformation process. Initially, the router discards layer 2 information, preparing for determining the packet's destination. This determination is made by referencing the CEF table, or FIB, which guides the router in making a forwarding decision.

After determining the forwarding path, it directs the router to a specific entry in the adjacency table. The router retrieves the Layer 2 rewrite string from the Adjacency table, which enables the router to put a new Layer 2 header to the frame. Finally, the packet is switched out to the outgoing interface toward the next hop [3].

3 Overview of SD-WAN

SD-WAN falls under the broader category of software-defined networking (SDN). SDN employs a centralized approach to network management, separating the underlying network infrastructure from its applications. This decoupling of the data plane and control plane enables customers to centralize network intelligence, facilitating increased network automation, simplified operations, and centralized provisioning, monitoring, and troubleshooting. Cisco SD-WAN implements these SDN principles within the WAN context.

In the control plane, network policies such as routing and security are enforced by making decisions about how packets should be routed through specific routers and ports. On the other hand, the data plane focuses on packet forwarding, and contains minimal intelligence required to transfer or discard packets between ports within the same device.

The figure below illustrates applying SDN principles to the WAN are illustrated on Fig. 2.

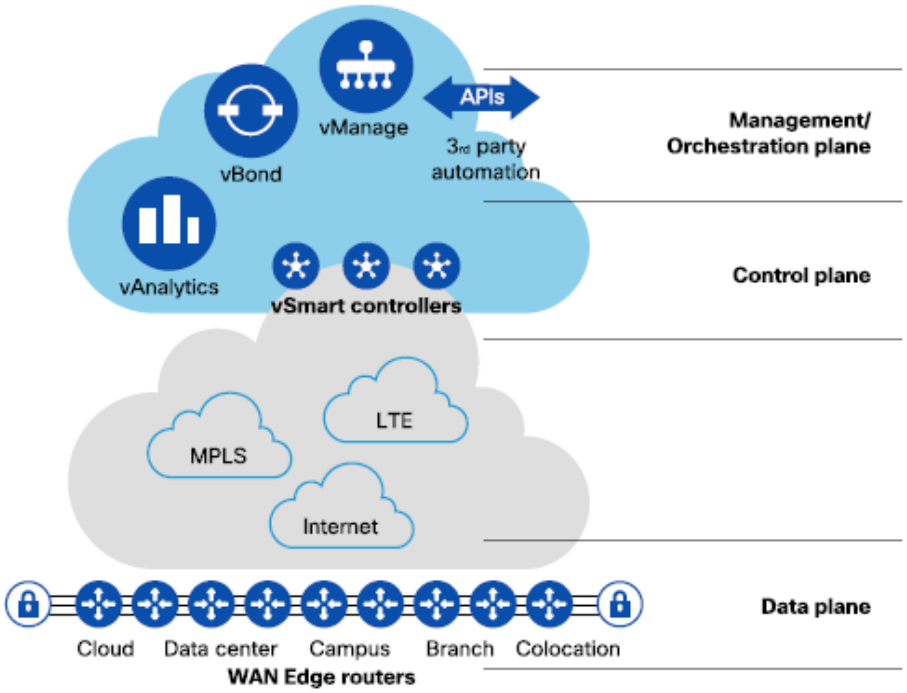


Fig. 2. SDN Principles in WAN [4].

The Cisco SD-WAN solution is made up of four planes: Data plane, Control plane, Management plane, and Orchestration plane. Below are described the main components for each plane.

Cisco vManage is the Management Plane of the SD-WAN system and serves as the centralized manager, handling all provisioning, configurations, monitoring dashboards, analytics, and maintenance for the entire SD-WAN network [5].

Cisco vSmart functions as the central control unit, managing topology building, traffic flow decisions, and control commands across the network. It supports fabric discovery, distributes policies to SD-WAN routers, and enforces centralized control plane policies [5].

Cisco vBond acts as the orchestrator within the SD-WAN system. This compoment conducts the initial authentication of WAN Edge devices and coordinates the connectivity among vSmart, vManage, and WAN Edge devices. It also plays a crucial role in facilitating communication between devices located behind Network Address Translation (NAT) [5].

Cisco WAN edge routers have only Data Plane or forwarding plane of the SD-WAN system. Edge routers are positioned at the boundaries of sites, such as remote offices, branch locations, campuses, and data centers. Their primary role involves establishing the network fabric and handles traffic forwarding, security, encryption, quality of service (QoS), routing protocols like Border Gateway Protocol (BGP) and Open Shortest Path First (OSPF), among other functions [5].

The overall data forwarding process is similar between the traditional WAN and SD-WAN. Cisco WAN edge routers can execute essential functions such as BGP, OSPF, Access Control Lists (ACLs), QoS, and various routing policies alongside overlay communication tasks. These routers ensure secure connectivity with all the control components and establish IPSec tunnels with other WAN edge routers to construct the SD-WAN overlay network. Furthermore, each WAN edge router establishes a control channel with every control element, facilitating the exchange of configuration, provisioning, and routing information [4].

4 Comparison of the Network Performance and the QoS between SD-WAN and the Traditional WAN

4.1 Research Methodology

For this research experiment, the network emulation software EVE-NG (Emulated Virtual Environment - Next Generation) is used as the primary tool. The version of the EVE-NG is 5.0.1-24 Community edition [6].

In EVE-NG, virtual machine images are used to create virtual network devices such as routers, switches, firewalls, and servers. These images can be used to create complex network topologies and test different network configurations [7].

The images that are imported in EVE-NG and the resources that are allocated for the devices for this experimental SD-WAN design are shown in Table 1.

Table 1. Images and resources for EVE-NG [6].

Device	Name of the image and version	Resources
vManage	vtmgmt-19.2.31	4 vCPU, 24 GB RAM
vSmart	vtsmart-19.2.31	2 vCPU, 2 GB RAM
vBond	vtbond-19.2.31	2 vCPU, 2 GB RAM
vEdge	vtedge-19.2.31	2 vCPU, 2 GB RAM
Windows server	winserver-S2012-R2-x64	12 GB RAM
Switches	I86bi_linux_l2-ipbasek9-ms.high_iron_aug9_2017b.bin	1 GB RAM
Internet router	L3-ADVENTERPRISEK9-M-16.4-2T.bin	1 GB RAM
MPLS router	L3-ADVENTERPRISEK9-M-16.4-2T.bin	1 GB RAM
NetEM	Linux-netem	4 GB RAM

Additionally, the tool NetEm is used to emulate packet loss because in the most cases the emulators cannot reach a loss that can be useful for analysis. NetEm is an open-source network emulator for Linux that can simulate various network conditions, including packet loss, latency, packet duplication, bursts, congestion, and packet re-ordering [8]. The NetEm tool in this case is connected between vEdge4 and the WAN transport Internet and Multi Protocol Label Switching (MPLS) as shown in the Fig. 3.

Network emulators play a crucial role for research experiments. They allow for the controlled testing of realistic network scenarios, something that cannot be accomplished using only real network devices without emulation features.

In this experimental SD-WAN design shown on figure 3, there is one controller site and four remote sites. The controllers vSmart, vManager and vBond together with CA server are in controller site. The WAN Edge routers together with the layer 2 switches and PCs are in the remote sites The remote sites are interconnected in a fully meshed topology. Each of these sites has internet connectivity. The WAN transport used are one Internet service provider and one MPLS.

The core switch is used for connecting the controllers together and to transport. The interfaces of the core switch that are connected to the controllers are members of VLAN 1. The Interface VLAN 1 is configured with an IP address that is used as gateway of the controllers and CA server. The interface of the

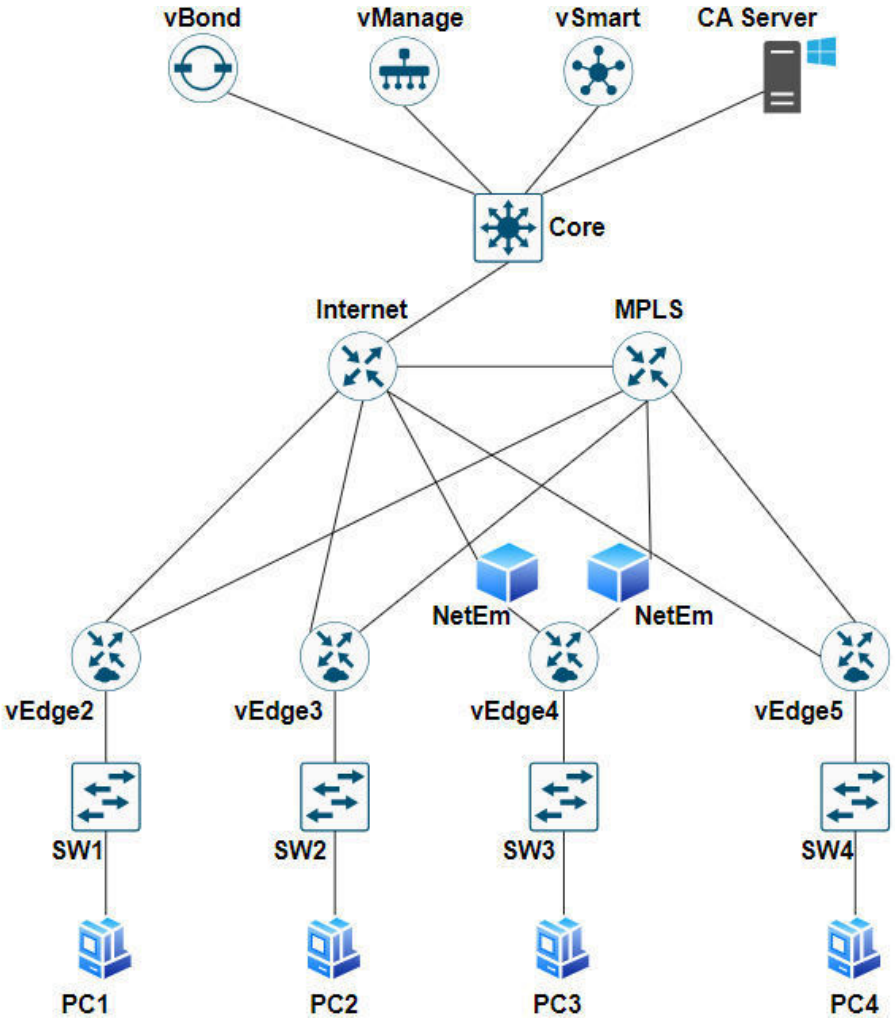


Fig. 3. Experimental Network Topology Design.

switch that is used for connection to the internet transport is routed port. There is also a default route configured with a next hop of the IP address of the internet router that is connected to this core switch. The transport site of the controllers is configured with VPN 0. The default route is also configured on the controllers with the next hop of the IP address of VLAN 1.

The CA server is an Active directory server that is promoted to a Domain controller and configured as a Certification Authority (CA) server. The CA certificates need to be installed in the controllers to authenticate each other. The CA server certificate, or the root certificate is imported to Cisco vManage. The Cisco vManage needs a certificate because vManage authenticates itself to vBond, and vSmart. The vBond and vSmart are added to vManage, and a Certificate Signing Request (CSR) of the vBond and vManage is generated. The CSR of the controllers are added to CA server and a signed CSR is given a certificate for vSmart and vBond that is imported to vManage. The root certificate is also installed on the vEdge's, and afterwards vEdges are registered to vManager.

The WAN edges are connected to both Internet and MPLS transports. A static default route pointing to the next hop gateway is configured for tunnel establishment on the Internet and MPLS transports. There is also VPN 0 configured as a transport connection with enabled tunnel interface and IPSec encapsulation [7].

The communication between WAN edge's and vSmart is via the control plane connection using the cisco proprietary protocol Overlay management protocol (OMP). The control plane connection is needed for sending and receiving control information such as routing updates, security information, policy information, etc. The security framework protocol in the control connection is Datagram Transport Layer Security (DTLS). Therefore, the traffic between WAN edges and controllers is over a secure connection. The communication between WAN edge's is via data plane connection. This type of connection is also secured using IPSec between the WAN edges or permanent tunnel between WAN edges.

The Internet router has two static routes. One static route for the network on the controller's site with next hop with the IP address of the interface on the switch that is connected to the internet router, and another static route for the network behind the MPLS router with next hop of the IP address of the interface on MPLS router that is connected to the Internet router.

The MPLS router has also two static routes. One static route for the network between the Internet router and controller's site and another static route for the network in the controller's site.

In this paper, FEC adaptive mode is configured for all applications but not for a specific application to address packet loss within the experimental SD-WAN design, utilizing the software network emulator EVE-NG.

It is configured a data policy that all traffic with any destination and any type of application should be used FEC adaptive if the fec-threshold for packet loss is above 1 %.

The analysis is conducted by randomly adding packet loss to the network flows before the packets are queued, accomplished through the utilization of the

network emulator NetEm. The percentage values (%) of packet loss that are specified for this analysis are 1 %, 2 %, 4 %, 8 % and 16 %. The following delay values (in milliseconds) were used: 5, 20, 30, 80 and 200. This packet loss and delay are correlated to the real-world values about these parameters while using the Internet and MPLS [9].

ICMP packets are used as a common method for packet loss testing by sending multiple pings from one branch location to another. Each of the packet loss levels were tested for 1000 ICMP packets that were sent to vEdge4 from the other vEdge's.

Moreover, following the completion of each test, the outcomes were thoroughly examined. In the event of any discrepancies detected, the test case was re-executed.

4.2 Analysis of Packet Loss

Generally, packet losses commonly arise from various factors including network congestion, overloaded devices, and issues with network hardware. Higher packet loss rates or increased latency can intensify delays, amplifying the impact of packets failing to reach their intended destinations.

IP is a fundamental network layer protocol within the TCP/IP suite. It operates as a connectionless protocol on the internet, offering no assurances regarding packet delivery or performance metrics like latency, jitter, packet loss, or bandwidth. Consequently, it is easier to encounter congestion or network degradation, which can compromise the QoS. Real-time applications like voice and video are affected more severely by packet loss. Therefore, in situations where packet loss is higher, or the media quality is essential, Forward Error Correction (FEC) can be used for recovering packet losses. The FEC feature allows critical applications to work well over unreliable WAN links, usually Internet circuits.

Cisco SD-WAN incorporates Forward Error Correction (FEC), which enables packet recovery at the layer 3 (IP) level. Implementing FEC at layer 3 within the network reduces the duration required for TCP to retransmit a dropped packet. Layer 3 FEC offers the advantage of being cognizant of other network traffic traversing the same link, allowing for intelligent management of all flows, whereas TCP is limited to awareness of individual flows [10].

FEC functions as a method for recovering lost packets on a link by sending extra "parity" packets alongside each set of four packets. Generating this parity packet can be achieved through various methods. One of the methods involves using exclusive OR (XOR) operations. XOR is a bitwise operator that performs logical operations. When the input bits are identical, the output is false (0); otherwise, it's true (1).

Consequently, if the receiver successfully receives all packets, the parity packet becomes redundant and is discarded. However, if any of the data packets are missing, the lost packet can be reconstructed by conducting the same bitwise exclusive OR operation on the received three packets alongside the parity packet. This obviates the necessity for the sender to resend the packet, thereby the packet loss is avoided [11-12].

The FEC process is shown on Fig. 4.

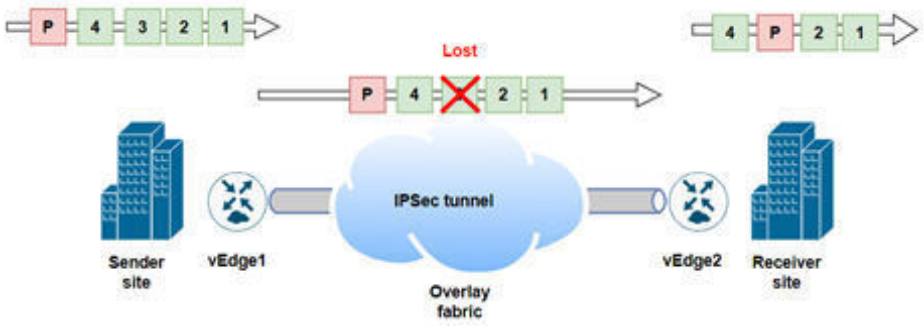


Fig. 4. Forwarding Error Correction Overview.

The FEC feature protects applications against experiencing packet loss along the transient network paths [13].

Using the FEC feature can increase the bandwidth consumption by 25 % because in each 4 packets 1 more packet is sent as a parity packet. 1 parity packet is as large as 1 data packet. With this more bandwidth we can protect the applications from packet loss. As a result, FEC can be configured in two distinct modes: FEC always and FEC adaptive. In the FEC always mode, a single parity packet is transmitted across the WAN link for every group of four packets at any given time.

In FEC adaptive mode, operation occurs only when the loss percentage of the transport link exceeds the configured FEC threshold value. This threshold, ranging between 1 % and 5 % packet loss, determines the activation point for FEC. For instance, setting it to 2 % means that once the Cisco SD-WAN fabric detects packet loss surpassing this threshold, FEC is automatically engaged. Contrarily, when the link loss diminishes below 2 %, FEC usage becomes unnecessary.

FEC is primarily utilized for loss-sensitive traffic such as VoIP, as even a slight increase in packet loss can significantly impact the entire communication process.

4.3 Mathematical Model

In traditional WAN packet loss recovery mechanisms typically rely on retransmission-based error recovery as a the most used transport layer approach to recover packet errors and losses. This mechanism involves the retransmission of lost packets upon the detection of errors.

Errors in packet transmission can be identified if the sender receives either a duplicate acknowledgment or a negative acknowledgment due to a missing sequence number. To provide feedback, the receiver sends an acknowledgment upon successful delivery of data. If any data is missing or received with errors, the

receiver sends either duplicate acknowledgments or negative acknowledgments. Upon receiving a negative acknowledgment or multiple duplicate acknowledgments, the sender initiates retransmission of the affected data [14].

Retransmission mechanisms are usually rejected in real-time audio and video streaming over the Internet due to their inherent retransmission delays. These delays can cause video packets to miss their intended playout time. When a packet needs to be retransmitted, it arrives at least three times the round-trip time after the transmission of the original packet. This delay may exceed the strict delay requirements of real-time audio and video streaming applications, where acceptable delay times typically range between 100ms and 200ms. Consequently, any retransmission of lost packets must occur within 100ms after the data loss to maintain the integrity of the audio and video stream [15].

The basic principle of Forward Error Correction (FEC) involves enhancing a communication channel with known error probabilities by using error-correcting codes to add redundant packets, known as parity packets. This allows the receivers to detect errors or losses and reconstruct the missing packets from the redundant ones, eliminating the need to retransmit the lost packets from the sender.

FEC can be implemented at both the bit level and the packet level. Typically, packet-level FEC is used for end-to-end communication, while bit-level FEC is applied to specific links. With bit-level FEC, redundancy bits are added to each sent packet, increasing its size. In contrast, packet-level FEC involves adding redundancy packets as separate entities within each transmitted block (group of packets). Consequently, the transmitted block consists of both data packets and redundancy packets [15-17].

Compared to retransmission mechanisms, FEC is recognized as an effective method for packet loss recovery. FEC is more suitable for real-time audio and video streaming because it introduces minimal end-to-end delay and ensures reliable transmission by incorporating extra redundancy information.

Adding more redundancy packets than necessary can consume more bandwidth and potentially overload the network during heavy traffic conditions. But, on the other hand, adding fewer redundancy packets than needed will reduce overall network load but may lead to insufficient recovery of lost packets in scenarios with high packet loss rates.

The packet loss is often unpredictable and fluctuates over time, the amount of extra data included with the original data in an FEC block is determined at the beginning of a session. This redundancy is calculated based on a long-term average of network losses.

$$R = \frac{n-k}{k} \tag{1}$$

The efficiency and effectiveness of FEC in recovering from packet loss depends on the redundancy factor R, defined as the number of parity packets (n-k) added to a block of k data packets, as described in equation 1 [18].

The aim of adaptive FEC is to dynamically adjust the level of redundancy based on the network's condition. It decreases the amount of redundancy when

packet loss at the receiver is low and increases redundancy when the network is congested, thereby minimizing the overhead caused by FEC [19].

The Gilbert-Model is a commonly used approximation for network loss, employing a Two-state Markov chain to represent end-to-end packet loss. This model is favored for its simplicity and mathematical clarity. In the Two-state Markov model, the "0" state signifies a lost packet, while the "1" state indicates successful packet delivery.

Let p represent the transition from state 0 to state 1, and q represent the transition from state 1 to state 0. Thus, the probability of packet loss is (1-q), and the probability of losing n consecutive packets is (1-q)$q$$n$-1. From the Markov chain transition matrix, the long-run loss probability π can be defined as:

$$\pi = \frac{p}{p+q} \tag{2}$$

The FEC adaptive predicts network losses and employs probability equations to estimate the expected loss at the receiver. Since each packet can only be either delivered or lost, the situation can be modeled with a binomial distribution.

In general, if we consider the loss of a packet (l) as a success event, where l follows a binomial distribution with parameters n (the number of packets) and pi (the probability of one packet being lost from qn packets), we write $l \sim$ B (n, π) [18],[20].

The probability of one packet being lost from a group of n packets is given by:

$$P(l) = \binom{n}{k} * (1-\pi)^{\text{n-1}} * \pi^{\text{n}} \tag{3}$$

Network packet loss exhibits significant fluctuations over time, with long-term averages often failing to accurately reflect the current network conditions. The proposed AFEC mechanism addresses this issue by estimating the expected channel loss probability using packet loss information received from the receiver. The estimated block loss probability for a group of k lost packets is calculated as follows:

$$E(x) = \sum_{l=0}^{k-1} \binom{n}{l} * (1-\pi)^{\text{n-}l} * \pi^{l} \tag{4}$$

The calculated average block loss probability is:

$$E(x) = \sum_{l=0}^{k-1} \binom{n}{l} * (1-\pi)^{\text{n-}l} * \pi^{l} * \frac{n-l}{l} \tag{5}$$

where:

E(x) is the estimated packet loss probability without FEC,

n is the total number of sources and parity packets per block,

k is the number of source packets per block, and

π is the channel loss probability [21].

4.4 Discussion of the Results

The findings reveal that there are reconstructed packets across all levels of emulated packet loss, signifying an enhancement in reducing the packet loss percentage and delivering improved Quality of Service (QoS) within Cisco SD-WAN in contrast to Traditional WAN. Packet loss is quantified as a percentage, calculated by comparing the total number of packets sent against those received:

$$Packet\ Loss\ \% = (\frac{Packets\ Sent\ -\ Packets\ Received}{Packets\ Sent}) * 100\ \% \tag{6}$$

where:

Packets Sent is the total number of packets being transmitted,

Packets Received is the total number of packets being successfully received.

Table 2. FEC Packet Loss Measurement.

Device	vEdge3 to vEdge4	vEdge3 to vEdge4	vEdge5 to vEdge4	vEdge2 to vEdge4	vEdge5 to vEdge4
Emulated packet loss (%)	1	2	4	8	16
Sent Packets	1000	1000	1000	1000	1001
Received packets	984	986	957	894	779
Sent parity packets	249	249	247	249	251
Reconstructed packets	16	12	37	53	65
Packets lost	16	14	43	106	222
Reconstructed packets (%)	100	86	86	50	29

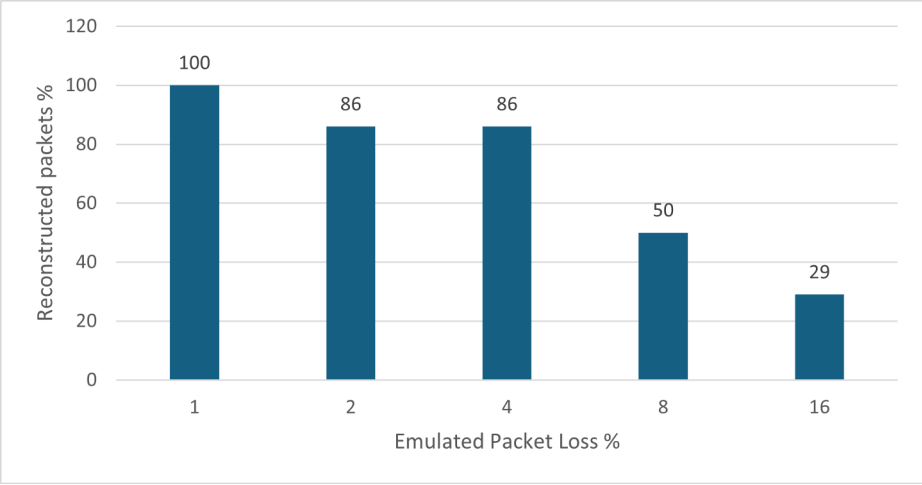


Fig. 5. Percentage graphic of packet loss reconstruction success rate.

The results of the experimental performance tests using FEC feature regarding packet loss are shown in the Table 2.

The data from the table above is crucial for evaluating the performance of the SD-WAN and its ability to maintain data integrity under varying conditions of packet loss. The parity packets play a vital role in reconstructing the lost packets, which is essential for ensuring reliable communication across the network. The decreasing trend in the percentage of reconstructed packets as the emulated packet loss increases indicates the limits of the current reconstruction mechanism under stress.

The reconstruction success rate using FEC is given on Fig. 5. It can be noticed that as emulated packet loss is increased the rate of the reconstructed packets is also decreased. In particular if the emulated packet loss is about 8 % the rate of the reconstructed packets drops to 50 %.

Higher packet loss values often result in multiple packet losses simultaneously, rendering them unrecoverable by FEC. Consequently, FEC is unable to achieve 100 % packet loss reconstruction.

4.5 Latency and Jitter

Cisco SD-WAN cannot directly reduce jitter and latency, as these parameters are inherent characteristics of the network path. However, it can indirectly improve them by reducing packet loss, which often leads to better jitter and latency performance.

Additionally, Cisco SD-WAN can achieve lower latency through intelligent routing and optimization features. It can also mitigate jitter by prioritizing traffic and implementing Quality of Service (QoS) mechanisms.

When packet loss occurs, the network may need to resend packets, resulting in varying delays. This leads to increased latency, as the time it takes for data to reach its destination grows. Additionally, packet loss can cause packets to arrive out of order, requiring extra processing time for devices to reorder them, further contributing to latency.

To compensate for packet loss or out-of-order packets, devices may buffer data, which adds variations in packet delivery times and increases jitter. Jitter, in turn, might indicate worsening congestion before it escalates into packet loss [23]. By reducing packet loss, packets can travel through the network more smoothly, minimizing delivery time fluctuations, which improves both jitter and latency.

Jitter measures the variation in packet arrival times [24] and can be formulated in the following equation:

$$Jitter_{(delay\ var)} = \frac{\sum variation\ delay}{\sum packet\ received} \tag{7}$$

Latency (or delay) measures the time it takes for data packets to travel from one point to another, representing the transmission delay that occurs as data moves toward its destination [25]. It can be defined with the following equation:

$$RTT = t_{returned} - t_{sent} \tag{8}$$

where:

RTT (Round Trip Time) is the duration it takes for a packet to travel from the source to the destination and then return to the source,

t_{sent} is the time when the packet is transmitted,

$t_{returned}$ is the time when the acknowledgement for that packet is received.

One-way latency is generally estimated to be half of the Round-Trip Time (RTT) if the network paths in both directions are symmetrical:

$$One\text{-}way\ Latency = \frac{RTT}{2} \tag{9}$$

Cisco SD-WAN continuously measures latency, jitter and packet loss using methods such as IP-SLA probes to make intelligent decisions about routing traffic. By doing so, they can select the optimal path and prioritize critical data, ensuring better overall network performance and reliability, minimizing latency and, consequently, jitter [26].

Different applications have varying requirements for traffic latency and jitter. For instance, Voice over IP (VoIP) services are particularly sensitive to these factors, as high latency and jitter can greatly diminish voice quality, rendering the service unacceptable to users. VoIP typically tolerates delays of up to 150 ms before call quality degrades to an unacceptable level. In contrast, data transfer is generally more tolerant of delays and jitter [27 - 28].

5 Conclusion

According to the packet loss analysis during the transmission it can be concluded that SD-WAN feature FEC can reduce packet loss and improve reliability by reconstructing the lost packets. FEC feature can mitigate the effects of packet loss by adding redundant packets to the sender to reconstruct lost packets. The experimental results with FEC indicated that a higher percentage of packets were successfully reconstructed with lower values of packet loss. However, as packet loss increases, the number of reconstructed packets decreases. Therefore, FEC cannot achieve 100% packet loss reconstruction. FEC feature in Cisco SD-WAN are particularly beneficial for some applications that cannot tolerate data loss, such as VoIP, video conferencing, real-time analytics, emergency services, financial transactions, etc.

Most applications consider a low packet loss level (1 %-2 %) acceptable. However, a packet loss rate of 5 % or higher can notably affect both network performance and user experience [22].

Based on the analysis and findings presented in this paper, future research directions for Cisco SD-WAN could focus on developing more efficient Forward Error Correction (FEC) algorithms. These algorithms should aim to manage higher levels of packet loss while minimizing performance degradation.

References

1. Segeč, P., Moravčik M., Uratmová J., Papán, J., Yeremenko, O.: SD-WAN - architecture, functions and benefits. In: 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), pp. 593–599. Košice, Slovenia (2020). https://doi.org/10.1109/ICETA51985.2020.9379257
2. Edgeworth, B. et al.: CCNP and CCIE Enterprise Core ENCOR 300-401., pp. 25–28 (2019).
3. Sherla, S.: INTRODUCTION TO CISCO EXPRESS FORWARDING (CEF).: https://blog.octanetworks.com/introduction-to-cisco-express-forwarding-cef/, last accessed: 2024/03/28.
4. Cisco.:Cisco SD-WAN Cloud Scale Architecture. pp. 20–22. : https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/sd-wan/nb-06-cisco-sd-wan-ebook-cte-en.pdf, last accessed: 2024/04/05.
5. Cisco.: Cisco SD-WAN Design Guide, https://www.cisco.com/c/en/us/td/docs/solutions/CVD/SDWAN/cisco-sdwan-design-guide.html, last accessed: 2024/06/15.
6. EVE-NG Ltd, https://www.eve-ng.net/, last accessed: 2024/03/15.
7. Harahus, M., Čavojský, M., Bugár, G., Pleva, M.: Interactive Network Learning: An Assessment of EVE-NG Platform in Educational Settings. Acta Electrotechnica et Informatica, Vol. 23, No. 3, pp. 3–9,(2023), https://doi.org/ 10.2478/aei-2023-0011.
8. Channa, J.: (2021, March 2). Emulating network latency and packet loss in Linux. (2021), https://www.jagchanna.ca/emulating-network-latency-and-packet-loss-in-linux/ , last accessed: 2024/05/15.
9. NetworkAcademy.io.: Packet loss, Latency and Jitter on EVE-NG. (no date), https://www.networkacademy.io/ccie-enterprise/sdwan/packet-loss-latency-and-jitter-on-eve-ng, last accessed: 2024/04/23.
10. Chatterjee, A.: Cisco Application Quality of Experience (APPQOE) for WAN optimization. (2022), https://community.cisco.com/t5/networking-blogs/cisco-application-quality-of-experience-appqoe-for-wan/ba-p/4532204, last accessed: 2024/03/25.
11. Forward error correction – Does your network need it?. (2016), https://talkingpointz.com/forward-error-correction-does-your-network-need-it/, last accessed: 2024/04/09.
12. Cisco.: Forward Error Correction, https://www.cisco.com, last accessed: 2024/04/28.
13. NetworkAcademy.io.: LAB 6 - Forward Error Correction (FEC). (no date), https://www.networkacademy.io/ccie-enterprise/sdwan/forward-error-correction-fec, last accessed: 2024/04/17.
14. Wang, -H. C., Chang, -I. R., Ho, -M. J., Hsu -C. S.: Rate-sensitive ARQ for real-time video streaming. IEEE Global Telecommunications Conference, GLOBECOM '03, vol. 6, pp. 3361-–3365, (2003).
15. Wu, D., Hou, T. Y., Zhang, -Q. Y.: Transporting real-time video over the Internet: challenges and approaches. Proceedings of the IEEE, vol. 88, no. 12, pp. 1855-–1877, (2000).
16. Basalamah, A., Sato, T.: A Comparison of Packet-Level and Byte-Level Reliable FEC Multicast Protocols for WLANs. IEEE Global Telecommunications Conference, GLOBECOM '07, pp. 4702-–4707, (2007).
17. Li, Z., Khisti, A., Girod, B.: Forward Error Protection for low-delay packet video. 18th International Packet Video (PV) Workshop, pp. 1-–8 (2010).

18. AL-Rousan, M., Nawasrah, A.: Adaptive FEC Technique for Multimedia Applications Over the Internet. Journal of Emerging Technologies in Web Intelligence, vol. 4, no. 2, pp. 142–147 (2012). https://doi.org/10.4304/jetwi.4.2.142-147
19. Kwon, -W. Y., Chang, H., Kim, J.: Adaptive FEC control for reliable high-speed UDP-Based media transport. Proceedings of the 5th Pacific Rim Conference on Advances in Multimedia Information Processing - Volume Part II, , pp. 364—372, Berlin, Heidelberg, (2004).
20. Nonnenmacher, J., Biersack, W. E., Towsley, D.: Parity-based loss recovery for reliable multicast transmissio. IEEE/ACM Transactions on Networking, vol. 6, no. 4, pp. 349—361, (1998).
21. Baldantoni, L., Lundqvist, H., Karlsson, G.: Adaptive end-to-end FEC for improving TCP performance over wireless links. 2004 IEEE International Conference on Communications, vol. 7, pp. 4023—4027, (2004).
22. Lamberti, A.: What is packet loss and how does it affect network performance?, https://medium.com/obkio/what-is-packet-loss-how-does-it-affect-network-performance-c38fe32ce2e7, last accessed: 2024/04/10.
23. Stallings, W.: Data and Computer Communications, 10th edition, Pearson Education (2014).
24. Sugeng, W., Istiyanto, J. E., Mustofa, K., and Ashari, A.: The Impact of QoS Changes towards Network Performance. In: International Journal of Computer Networks and Communications Security, vol. 2, no. 3, pp. 48–53, (2015).
25. Fahlborg, D: Measuring one-way Packet Delay in a Radio Network, Master of Science Thesis in Electrical Engineering Department of Electrical Engineering, Linköping University, Sweden (2018).
26. Cisco.: Measuring Delay, Jitter, and Packet Loss with Cisco IOS SAA and RTTMON, https://www.cisco.com/c/en/us/support/docs/availability/high-availability/24121-saa.html, last accessed: 2024/04/23.
27. Manova, R. Y., Sukmadirana, E., and Nurmanah, N. S.: Comparative Analysis of Quality of Service and Performance of MPLS, EoIP and SD-WAN. In: 2022 1st International Conference on Information System and Information Technology (ICISIT), Yogyakarta, Indonesia, pp. 403–408, (2022). https://doi.org/10.1109/ICISIT54091.2022.9872806.
28. Manova, R. Y., Sukmadirana, E., and Nurmanah, N. S.: An SD-WAN solution assuring business quality VoIP communication for home based employees. In: 2019 International Conference on Smart Applications, Communications and Networking (SmartNets), Sharm El Sheikh, Egypt, pp. 1–6, (2019). https://doi.org/10.1109/SmartNets48225.2019.9069755.

AI-Driven Security Mechanisms in Android: Securing Mobile Applications

Mila Dodevska [1*], Marija Taneska [1†] and Slave Temkov [1†]

[1]Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Rugjer Boshkovikj 16, Skopje, 1000, North Macedonia.

*Corresponding author(s). E-mail(s): mila.dodevska@finki.ukim.mk;
Contributing authors: marija.taneska@finki.ukim.mk;
slave.temkov@finki.ukim.mk;
†These authors contributed equally to this work.

Abstract

This paper gives an overview of how artificial intelligence (AI) is boosting Android application security, with a focus on the upcoming Android 15 release this fall. It explores AI's role in detecting threats, conducting real-time analysis, and implementing dynamic access controls. Additionally, it examines recent improvements in Android security designed to combat fraud, scams, and privacy breaches. The study underscores the collaborative efforts among Android, OEMs (Original Equipment Manufacturers), and developers in enhancing security. As cyber threats continue to evolve, AI-driven security mechanisms offer promising solutions for protecting electronic information and ensuring the integrity of mobile applications. Ultimately, it emphasizes the significance of proactive security measures and teamwork in protecting user privacy and upholding the trustworthiness of Android applications.

Keywords: Android, AI-Driven Security, Mobile Applications, Threat Detection, Machine Learning (ML) in Cybersecurity, Vulnerability Management, Data Privacy

1 Introduction

The role played by artificial intelligence in enhancing Android application security involves examining the evolution of security measures within the Android ecosystem. The combination of AI and Android app security is a groundbreaking development that

transforms the landscape of digital protection. In previous times, computer security has mainly depended on static rules and signature-based detection techniques. However, these methods have often been found unable to adapt to new incoming threats since they stick to certain weaknesses. [1, 2]

AI has revolutionized the security space, and things have become highly dynamic and aggressive. The use of AI-powered solutions employs machine learning algorithms and advanced data analytics methods that are capable of detecting and responding to threats in real-time, enhancing the resilience of Android applications against cyber-attacks. [3] This transition from reactive to proactive security is a major milestone toward safeguarding user data confidentiality and upholding mobile applications consistency in an expanding digital world. [2, 3]

As cyber threats are becoming more sophisticated, the integration of AI-driven security is required in Android applications at this stage. Such AI-powered systems are designed to change as cyber threat environments change, learn from patterns, and anticipate possible flaws even before they can be exploited. By doing so, AI-driven helps prevent from causing serious harm to users' trust or safety, even with an increase in online activities. [2] This research paper aims to discuss the importance of AI in strengthening Android applications with a focus on various technologies and tactics that have redefined mobile security systems in the digital age.

This paper explores the pivotal role of AI in transforming Android application security, beginning with a comprehensive overview of traditional security measures and their evolution. Section 2 presents related work, offering a review of key studies that have shaped the field of AI-driven security in Android applications. Section 3 analyses the different aspects of Android Application Security, examining AI's contribution to intelligent threat detection and prevention (3.1) and real-time analysis (3.2). The analysis extends to the importance of real-time security updates (3.3) and innovative methods like behavior-based authentication (3.4) that enhance app security. The paper further explores dynamic access control mechanisms (3.5) and proactive vulnerability management strategies (3.6), concluding with a look at advanced encryption techniques (3.7) that fortify data protection. The exploration continues in Section 4 with an analysis of Google Play Protect, highlighting its role in combating fraud and scams (4.1) and shielding against screen-sharing social engineering tactics (4.2). This section also covers next-generation cellular security (4.3) and the empowerment of both developers and users through enhanced protective measures (4.4). The next section is a short discussion on the findings and proposes future research directions in this field (5). Finally, the paper concludes with a synthesis of the key findings and their implications for the future of Android application security (6).

2 Related work

The research landscape on AI-driven security mechanisms in Android applications has been significantly enriched by several key studies that explore various dimensions of cybersecurity, dynamic analysis, and database security.

AI Sentry: Reinventing Cybersecurity Through Intelligent Threat Detection [1] leverages AI to detect and prevent threats in real-time. By employing machine learning and neural networks, the system can identify anomalies and predict zero-day attacks. This innovative approach aligns with our research goal of enhancing Android application security against sophisticated cyber threats.

Dynamic Security Analysis on Android: A Systematic Literature Review [2] presents a comprehensive overview of techniques used to analyze Android systems in real-time. The paper focuses on network monitoring, system-call tracing, and taint analysis, identifying research gaps and limitations in automated testing tools. This review informs our research on real-time security updates and dynamic access control in Android applications, providing valuable insights into the challenges and opportunities in this field.

Another important study on this topic is by Alkahtani & Aldhyani, which concludes the SVM and LSTM models are the most effective in detecting Android malware and outperformed other state-of-the-art models. [3]

AI Techniques for Software Vulnerability Detection and Mitigation is one of the foundational works in this area and [4] discusses how AI can be used to identify and address software vulnerabilities. This research aligns with our focus on proactive vulnerability management and the use of advanced AI techniques to enhance the security of Android applications.

The Performance-Sensitive Malware Detection System Using Deep Learning on Mobile Devices by Feng et al., suggests a proactive malware detection solution MobiTive, that can run directly on mobile devices and provides detection accuracy of different deep neural networks and real-time detection performance. [5]

These studies collectively provide a strong foundation for the exploration of AI-driven security mechanisms in Android applications. By integrating the insights gained from these works, our research aims to further advance the field by proposing innovative solutions that enhance the resilience of Android applications against emerging cyber threats.

3 Android Application Security

The future of Android application security [6] will be powered mainly by Artificial Intelligence (AI) [7]. Neural Networks and Machine Learning Algorithms, as mentioned by [3, 8] come in handy as governing principles behind intelligent risk detection and prevention. AI-powered, analyzed rapid trends and behaviors, proactive defenses that detect and reduce potential threats before they grow to breaches. In this way, it shall improve the safety of the Android applications and give confidence to the users, and also preserve the privacy of their data.

3.1 The pivotal role of Artificial Intelligence in intelligent threat detection and prevention

Advancement in Android application security, largely fostered by Artificial Intelligence (AI), has led to a considerable improvement in intelligent prevention and detection of risks [1]. These AI-based systems differ from the old defense measures because

they can examine trends as well as behaviors that happen very fast and detect abnormal activities faster than any other technology. The latest machine learning techniques, including supervised, unsupervised, and ensemble methods, have demonstrated significant potential in improving mobile attack detection and prevention [9].

There is an active defense system put up by AI algorithms that identify even suspicious activities long before they become breaches hence this algorithm makes it possible to detect breaches before they actually occur. This method reduces the chances of ill intentions, thus making Android applications safe for users. According to [8], another aspect that comes in handy is the artificial neural networks which are used to simulate neurons and process large amounts of data to detect intrusions, secure mobile applications, and protect privacy.

3.2 Real-time Analysis

For the security of applications, real-time analysis is fundamental. It enables application behavior, network traffic, and system interactions to be continuously monitored for any threats that may arise. This makes it possible for AI systems to respond fast to any suspicious happenings thus weakening risks that could have led to a breach in security or making an application environment safer.

In contrast with conventional methodologies, such an approach differentiates AI-based security solutions from traditional techniques, as they can themselves against emerging threats and zero-day vulnerabilities. Real-time analysis also leads to low rates of false positives since it takes into account contextual factors such as user behavior that enhance its detection abilities [2]. However, integration with incident response procedures allows for quick mitigation measures, reducing the potential impact of information violations upon users and businesses.

3.3 Real-time Security Updates

AI makes real-time security updates possible, which is crucial for safeguarding applications from constantly changing cyber threats. AI instantaneously monitors and analyzes the behaviors of applications for rapid detection of new threats and vulnerabilities [10]. This would prepare the developers to quickly issue updates that would fix these security issues proactively and effectively. This dynamic response registers confidence in users about their applications being properly fortified against new risks.

Real-time security updates, enabled by AI, are vital in combating evolving cyber threats. AI algorithms monitor application behavior and network activities, identifying potential security risks as they arise, more specifically about Play Integrity API and Google Play Protect is elaborated in 4.4. This proactive approach allows developers to release timely updates, patching vulnerabilities, and enhancing application security. By staying ahead of emerging threats, AI-driven security measures provide users with peace of mind, knowing their applications are continually protected against the latest risks.

3.4 Enhancing App Security through Behavior-Based Authentication

Authentication is fundamental to application security, and AI introduces behavior-based authentication as a dynamic approach. Behavior-based authentication, powered by AI, revolutionizes traditional methods by dynamically assessing user behavior patterns for identity verification [11]. This approach provides a more secure and user-friendly authentication experience, to individual user habits over time. By understanding typical user behaviors, AI ensures a seamless yet robust authentication process, improving overall application security [12]. Behavior-based approaches using machine learning techniques have shown promise in capturing and learning user behavior for anomaly detection on mobile devices [10] By continuously learning and adjusting, AI-driven authentication enhances the overall security posture of applications while ensuring a seamless login process for users.

3.5 Dynamic Access Control

Imagine an AI-powered access control system acting as an intelligent patrol for your data, much like a highly perceptive bodyguard. Unlike traditional rule-based systems that are clumsy and vulnerable to manipulation, this advanced system finely tunes its responses based on your specific usage patterns. It takes into account contextual factors such as your location, current activities, and frequency of application usage to dynamically determine access levels. This ensures that while your applications remain tightly secured, your user experience remains seamless. Furthermore, the AI's capabilities allow it to continuously evolve in response to emerging threats, thereby enhancing the protection of your information.

The integration of AI into access control mechanisms within Android applications represents a significant advancement over traditional static rule-based systems, which are often vulnerable to exploitation. AI-driven access controls, in contrast, dynamically adjust permissions in response to real-time user interactions, device context, and potential risk factors [13]. This capability allows for a more precise allocation of access rights, enhancing security while maintaining a seamless user experience.

3.6 Proactive Vulnerability Management

Imagine a paradigm in which security within applications surpasses the constant reactive cycle of countering hacker advances—this is the potential of AI-driven predictive mechanisms in Android application development. Rather than relying on inflexible rules that hackers can easily evade, AI systems leverage historical exploit data, user behavior patterns, and emerging threat indicators to anticipate vulnerabilities before they are exploited. This approach functions are similar to a forward-looking security consultant, predicting and preempting cyber threats in an ongoing strategic engagement. This transition from a reactive to a proactive security model promises a significantly more secure and user-centric experience, fundamentally enhancing the landscape of application security. [2]

The world of Android applications is like a battlefield where hackers are constantly trying to find new ways to break into devices. What if we could predict how they would

attack? That's where AI-powered vulnerability management comes in. By analyzing vast amounts of data on past vulnerabilities, user interactions with applications, and real-time threat intelligence, AI can accurately predict potential security risks [13]. This provides developers with a crucial head start during which they can fix vulnerabilities and implement stricter access controls before attackers even know where to look. It greatly enhances the overall security of Android applications and makes the mobile environment more trustworthy for users.

3.7 Advanced Encryption Strategies

In today's world, we have been witnessing numerous data breaches. In such a world, encryption becomes very vital in ensuring sensitive information is well protected. AI, therefore, brings advanced techniques of encryption that are far from the conventional methods to help guarantee that data on Android security applications are secured and private. From end-to-end protocols to the robustness brought about by the implementation of imposing algorithms, AI-driven encryption thus assures intensified data protection and gives trust to the users in terms of application security [14].

AI-driven encryption techniques in the security of data within Android applications are important milestones. They enhance the conventional methods with advanced algorithms and self-adaptive capabilities for strengthening data confidentiality. With end-to-end encryption protocols in place and robust algorithms at work, AI ensures that sensitive information is still safe despite evolving threats. Improved encryption around user data not only helps in protection but also builds trust and confidence in the security measures taken within Android applications.

4 Google Play Protect: Ensure your device's security

Google Play Protect is a security element within the Google Play Store that acts like an automated attempt against malware applications and other security threats on Android devices. Operating continuously in the background, it scans for potential dangers to maintain device security. This research paper explores the functionalities of Google Play Protect to understand its role in securing the Android ecosystem.
Google Play Protect (GPP) examines 200 billion Android applications daily, safeguarding over 3 billion users from potential malware threats. To enhance fraud and abuse detection, Google is extending the capabilities of Play Protect with live threat detection, leveraging on-device AI. This feature enables Play Protect to analyze additional behavioral cues related to the utilization of sensitive permissions and interactions with various applications and services. If suspicious activities are identified, the application may undergo further scrutiny by Google, and users will be promptly warned, or the application disabled if malicious behavior is confirmed. [15]
The identification of suspicious behavior occurs on the device itself, ensuring user privacy through the Private Compute Core mechanism, which provides protection without compromising user data. Manufacturers such as Google Pixel, Honor, Lenovo, and others are set to integrate live threat detection into their devices later this year.

4.1 Fortifying Defenses: Strategies to Combat Fraud and Scams

In Android 15, significant efforts are being made to improve protections against fraud and scams, with a focus on enhancing user security and privacy. One important update is that one-time passwords (OTPs) will no longer be shown in notifications, except for certain apps like wearable companions. This change will help prevent fraudsters and spyware from accessing OTPs through notifications, making users safer from malicious attacks [16].

Additionally, Android 15 introduces expanded restricted settings that mandate additional user approval for application permissions when side-installing from Internet sources such as web browsers or messaging applications. This added security layer helps alleviate the misuse of sensitive permissions by fraudsters, reducing the risk of unauthorized access to personal data. [6]

Furthermore, ongoing advancements in AI-driven protections, such as scam call detection using on-device Gemini-Nano AI, underscore Android's proactive defense strategy against evolving threats. These initiatives highlight Android's commitment to providing a secure mobile environment, alleviating risks associated with fraud and scams, and ensuring users can trust their devices with their sensitive information.

4.2 Shielding Against Screen-Sharing Social Engineering Tactics

In Android 15, enhanced controls for screen sharing are being implemented to combat social engineering attacks aimed at accessing users' screens and stealing sensitive information. These measures include automatically hiding notifications and one-time passwords (OTPs) during screen-sharing sessions, preventing remote viewers from seeing confidential details. [16]

Moreover, Android 15 introduces safer login procedures that ensure users' screens are hidden when entering credentials like usernames, passwords, and credit card numbers during screen sharing. This added protection significantly reduces the risk of unauthorized access to personal information, strengthening user privacy and security. Additionally, Android devices will soon provide the option to selectively share content from individual applications rather than the entire screen, giving users greater control over their screen privacy.

Recognizing the importance of clear indicators for content sharing, Android is introducing a more prominent screen indicator to alert users when screen sharing is active. What that is going to do is that in the future, this will let people see how much of their data is exposed to view and enable them to leap quickly out of screen sharing with a single tap. This will let them have control of privacy under collaborative interactions. These developments drive home Android's commitment to improving user security and privacy in an advancing digital world.

4.3 Next-Generation Cellular Security: Fighting Fraud and Surveillance

In Android 15, a comprehensive suite of advanced mobile security measures has been deployed to counter increasing threats of fraud and surveillance, particularly through

the use of cell site simulators. Among these innovations, Cellular Cipher Transparency stands out as a critical defense mechanism. It alerts users when their cellular network connections are unencrypted, empowering them to recognize potential interception attempts by malicious actors. This proactive notification system enables users to take immediate steps to protect their sensitive communications from exploitation [16].

Additionally, Identifier Disclosure Transparency emerges as another crucial safeguard, especially for high-risk individuals such as journalists or activists. It notifies users of any unauthorized tracking of their locations through false cellular base stations or surveillance tools, providing essential awareness and enabling them to diminish potential threats effectively.

The deployment of these cutting-edge security measures requires close collaboration with device OEMs and the integration of compatible hardware components. This collaboration underlines Android's continuous efforts to drive a more secure mobile ecosystem and confirms that such developments shall be adopted incrementally in the coming years across devices. By focusing on such improvements, Android will foster better security and privacy for users, build trust in its platform, and arm them with stronger defenses against ever-changing digital threats.

In all, such proactive action taken towards the protection of its users by Android has not only led to added security for the individual user but has also made the digital space a little more difficult for actors desiring to exploit these flaws for malicious purposes. All of these efforts put Android at the very front rungs of protection of users' privacy and show its quest further to equip users with assigned safe mobile user experiences.

4.4 Empowering Developers and Protecting Users

Developers still face ongoing challenges in safeguarding their applications against scams and fraudulent activities, prompting the introduction of robust security measures. [16] A key addition is the Play Integrity API, an enhanced tool designed to verify application integrity and detect potential risks. This API allows developers to ensure their applications remain unmodified and are running on genuine Android devices. It enables the identification and prevention of fraudulent or risky behaviors by incorporating new in-application signals. These include checks for risks from screen capturing or remote access, detection of known malware through Google Play Protect, and identification of anomalous device activities. By leveraging these signals, developers can proactively secure sensitive applications like financial or banking applications, prompting users to take necessary actions such as closing risky applications or enabling Google Play Protect [15].

In addition to combating fraud and scams, Android 15 introduces upgraded policies and tools to strengthen user privacy, particularly concerning photo permissions. Starting August, applications on Google Play must justify their need for broad access to photo or video permissions, aiming to restrict unnecessary access to user data. Furthermore, Android's preferred photo picker solution now supports cloud storage services such as Google Photos, facilitating easier photo selection without requiring extensive permissions. Future enhancements will also enable local and cloud search

within the photo picker, enhancing user control over their data while maintaining convenience. These updates underline Android's commitment to empowering developers with advanced security tools and safeguarding user privacy in today's evolving digital landscape [6].

5 Discussion and future work

It is clear that AI has an increasingly significant role in enhancing Android application security, particularly in real-time threat detection, behavior-based authentication, and dynamic access control mechanisms. These technologies have shown effectiveness in reducing cyber threats and ensuring user safety. [1, 3]

However, challenges remain in continuously adapting these AI systems to new and unforeseen threats, which are increasingly sophisticated. Future research should focus on developing AI models that are not only robust against current threats but can also learn and adapt in real-time to emerging vulnerabilities. [2] Furthermore, balancing the enhanced security features with user convenience remains a critical issue, as overly strict measures can impact the user experience negatively.

Exploring the integration of AI with other emerging technologies, such as blockchain, could provide new avenues for improving security and privacy in Android applications. This integration could create a more secure, transparent, and decentralized framework, as suggested by Khater and Dawoud. [4]

Finally, the ethical implications of using AI for security purposes must be considered. Ensuring transparency, user privacy, and compliance with data protection regulations is essential to maintaining user trust. [13] Future work should also explore methods for making AI decisions explainable and understandable to users and developers alike, as well as simplifying the existing mechanisms of protection for mobile applications and devices.

6 Conclusion

Artificial intelligence integrated into Android application security, specifically in Android 15, is one giant leap toward protecting mobile applications. Needless to say, AI-driven solutions have outclassed the traditional ways of security through dynamic threat detection, real-time analysis, and proactive management of vulnerabilities. [1, 2] Proactive approaches that AI enables ensure timely identification and diminishing of prospective security threats. [3] Features like behavior-based authentication and dynamic access controls further enhance protection without compromising user experience. [4, 11]

Collaborative efforts between Android, OEMs, and developers underscore a shared commitment to application security and user privacy. Features such as Google Play Protect with live threat detection and stringent privacy policies exemplify this commitment to maintain user trust and ensure a secure mobile ecosystem. [6, 13]

Essentially, AI integration in the field of Android application security will be a milestone to increase the bar regarding mobile application safety and integrity. Android

offers proactive mechanisms of defense, constancy, real-time threat analysis, and strong protection against the ever-evolving nature of cyber threats with the help of AI-driven solutions. This strengthens the commitment of Android to a safer mobile experience along with powerful tools for developers to fight off fraud, scams, and other breaches in privacy effortlessly. [3, 8]

With continuous research on advancing the existing methods and exploring new ones, as well as addressing the challenges and enhancing the resilience and transparency of these solutions, we can ensure that Android applications remain secure and trustworthy, providing a robust defense against future cyber threats and protecting user data in an increasingly digital world.

References

[1] Rangaraju, S.: Ai sentry: Reinventing cybersecurity through intelligent threat detection. EPH-International Journal of Science And Engineering **9**(3), 30–35 (2023)

[2] Sutter, T., Kehrer, T., Rennhard, M., Tellenbach, B., Klein, J.: Dynamic security analysis on android: A systematic literature review. IEEE Access (2024)

[3] Alkahtani, H., Aldhyani, T.H.: Artificial intelligence algorithms for malware detection in android-operated mobile devices. Sensors **22**(6), 2268 (2022)

[4] Khater, H.M., Khayat, M., Alrabaee, S., Serhani, M.A., Barka, E., Sallabi, F.: Ai techniques for software vulnerability detection and mitigation. In: 2023 IEEE Conference on Dependable and Secure Computing (DSC), pp. 1–10 (2023). IEEE

[5] Feng, R., Chen, S., Xie, X., Meng, G., Lin, S.-W., Liu, Y.: A performance-sensitive malware detection system using deep learning on mobile devices. IEEE Transactions on Information Forensics and Security **16**, 1563–1578 (2020)

[6] Android: Secure Android devices documentation. https://source.android.com/docs/security (2024)

[7] IBM: What is artificial intelligence (ai)? (2024). https://www.ibm.com/topics/artificial-intelligence

[8] Bishtawi, T., Alzu'bi, R.: Cyber security of mobile applications using artificial intelligence. In: 2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), pp. 1–6 (2022). IEEE

[9] Al Hwaitat, A.K., Fakhouri, H.N., Alawida, M., Atoum, M.S., Abu-Salih, B., Salah, I.K., Al-Sharaeh, S., Alassaf, N.: Overview of mobile attack detection and prevention techniques using machine learning. International Journal of Interactive Mobile Technologies **18**(10) (2024)

[10] Rashad, S., Byrd, J.M.R.: Behavior-based security for mobile devices using

machine learning techniques. International Journal of Artificial Intelligence & Applications (2018)

[11] Progonov, D., Cherniakova, V., Kolesnichenko, P., Oliynyk, A.: Behavior-based user authentication on mobile devices in various usage contexts. EURASIP Journal on Information Security **2022**(1), 6 (2022)

[12] Jansen, W.A., Korolev, V.: A location-based mechanism for mobile device security. 2009 WRI World Congress on Computer Science and Information Engineering **1**, 99–104 (2009)

[13] Dawoud, A., Bugiel, S.: Bringing balance to the force: Dynamic analysis of the android application framework. Bringing balance to the force: dynamic analysis of the android application framework (2021)

[14] AI Dabagh, N.B., S Mahmood, M.: Multilevel database security for android using fast encryption methods. AL-Rafidain Journal of Computer Sciences and Mathematics **16**(1), 87–96 (2022)

[15] Google: Google play protect (2024). https://developers.google.com/android/play-protect/client-protections

[16] Titterington, A.: Security in Android 15: What's new? https://www.kaspersky.com/blog/android-15-new-security-and-privacy-features/51311/ (2024)

Integrating Cryptographic Techniques to Protect AI Systems and Data

John Ikechukwu[1], Azizakhon Toshpulatova[1], Elissa Mollakuqe[1][0000-0003-0508-105X] Ibrahim Isiaq Bolaji[1] Jose Luis Cano[1], Patrick Udeckukwu[1], Hasan Dag[1][0000-0001-6252-1870]

[1] Faculty of Management Information Systems, Kadir Has University, Istanbul, Turkey

cikechukwujohn@stu.khas.edu.tr
azizakhon.toshpulatova@stu.khas.edu.tr
elissamollakuqe@gmail.com
isiaqibrahim@stu.khas.edu.tr
canojl@stu.khas.edu.tr
patrik.udeckukwu@stu.khas.edu.tr
hasan.dag@khas.edu.tr

Abstract. The development of artificial intelligence (AI) technologies has led to profound changes in all industries, but has also raised significant security concerns. AI systems are vulnerable to various attacks and security vulnerabilities that can compromise sensitive data and system integrity. In this paper, we present a research initiative that focuses on securing AI systems through the application of cryptographic techniques. The main goal of this research is to address the urgent need for robust security measures in the field of AI. We aim to explore and implement cryptographic methods as a means to protect AI systems. In doing so, we emphasize their critical importance at a time when AI is increasingly integrated into numerous areas of our daily lives. Our research takes a multi-faceted approach. We begin with a comprehensive survey of the existing literature on AI security and cryptographic techniques. We analyze the vulnerabilities and risks associated with AI systems and explore how cryptographic methods can mitigate these threats. We select AI models relevant to healthcare and finance applications, using diagnostic models like convolutional neural networks (CNNs) for medical image analysis and fraud detection models such as decision trees or neural networks. Datasets include publicly available medical image databases (e.g., X-ray or MRI images) and credit card fraud detection datasets. We implement cryptographic methods including the Paillier cryptosystem for homomorphic encryption. Our preliminary results demonstrate the effectiveness of this techniques in protecting sensitive data, algorithms, and AI-generated content from unauthorized access and tampering. Our findings highlight the potential role of blockchain technology in transparency, traceability, and trustworthiness of AI-generated content. The significance of this research lies in its potential to revolutionize the AI security landscape. By using cryptographic techniques, AI experts can protect their systems against privacy breaches, adversarial attacks, and misuse of AI-generated content. The findings not only contribute to the theoretical understanding of AI security but also offer practical solutions that can be applied across industries, including healthcare,

finance, autonomous vehicles, and more. Ultimately, this research has the potential to increase public trust in AI technologies and foster innovation in a secure and reliable AI-driven world.

Keywords: *AI, Techniques, Homomorphic, Encryption, Privacy, Preservation Integrity and Trust*

1 Introduction

The advancement of artificial intelligence (AI) technologies has brought transformative changes across various industries, particularly in the healthcare sector. AI-driven diagnostic tools have significantly enhanced the accuracy and efficiency of medical image analysis, enabling early detection and treatment of numerous health conditions. However, this technological progress has also introduced substantial security concerns. AI systems are vulnerable to various attacks and security breaches, which can compromise sensitive patient data and the integrity of diagnostic processes [1]. In the context of healthcare, the protection of patient information is paramount. Regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe impose stringent requirements on data privacy and security [2]. Ensuring compliance with these regulations while leveraging AI technologies presents a significant challenge. Traditional encryption methods, while effective in protecting data at rest and in transit, are insufficient for securing data during processing by AI models. To address these challenges, our research focuses on integrating cryptographic techniques, specifically homomorphic encryption, with AI models used in healthcare diagnostics. Homomorphic encryption allows computations to be performed on encrypted data without needing to decrypt it first, thus ensuring that sensitive patient information remains protected throughout the diagnostic process [3]. This research aims to explore and implement homomorphic encryption methods to secure AI systems in healthcare, emphasizing their critical importance in maintaining data privacy and security. Our research methodology involves a multi-faceted approach, beginning with the encryption of medical image datasets using the Paillier cryptosystem [4]. We then utilize convolutional neural networks (CNNs) to analyze the encrypted images directly, evaluating the effectiveness of this approach in maintaining data confidentiality and achieving high diagnostic accuracy [5]. By doing so, we aim to provide a robust framework for secure and private medical image analysis, ensuring that patient information is protected even as it is processed by AI technologies. In our paper, we referred to the developed framework as "robust" to indicate its resilience against a variety of security threats and its ability to maintain patient privacy throughout the medical image analysis process. To clarify, we understand "robust" to encompass the following aspects: 1. **Comprehensive Security Measures** - Our framework employs multi-

ple cryptographic techniques, including homomorphic encryption and secure multiparty computation, to safeguard patient data against various vulnerabilities, including unauthorized access and data breaches. 2. **Adaptability to Evolving Threats** - The framework is designed to be flexible, allowing for the integration of emerging cryptographic methods and practices that can address new and evolving security challenges in AI and healthcare and 3. **Compliance with Regulatory Standards** - The framework adheres to stringent data protection regulations, such as HIPAA and GDPR, ensuring that patient privacy is upheld in accordance with legal requirements.

This research not only contributes to the theoretical understanding of AI security but also offers practical solutions that can be applied across various healthcare settings. By demonstrating the feasibility and effectiveness of integrating homomorphic encryption with AI models, we seek to enhance trust and confidence in AI-driven healthcare solutions, fostering innovation while ensuring compliance with data protection regulations.

2 Motivation

Our research highlights the significant benefits of integrating homomorphic encryption with AI-driven diagnostic models in healthcare. By encrypting medical image datasets, such as Chest X-rays and MRI scans, we ensure that sensitive patient information remains secure throughout the diagnostic process. This approach allows AI models, specifically convolutional neural networks (CNNs), to analyze encrypted data directly, achieving high accuracy rates in detecting medical anomalies while maintaining patient confidentiality. Implementing homomorphic encryption aligns with stringent data protection regulations like HIPAA and GDPR, ensuring legal compliance and enhancing data security. It also facilitates secure cross-institutional collaboration, allowing healthcare providers and research institutions to share encrypted datasets without compromising patient privacy, thus fostering scientific advancements and improving healthcare outcomes. By mitigating risks associated with insider threats and inadvertent data exposure, our approach builds trust and confidence in AI technologies and healthcare providers, encouraging patient participation in clinical trials and research initiatives. This research provides a robust framework for secure and private medical image analysis, demonstrating that it is possible to enhance diagnostic accuracy without compromising patient privacy. By using cryptographic techniques within AI models, we ensure the confidentiality and integrity of sensitive healthcare data, fostering innovation in a secure and reliable AI-driven healthcare landscape.

3 Literature review

Several cryptographic techniques have been developed and are being used to protect AI systems and data. These techniques aim to secure data during storage, processing, and transmission and ensure the integrity and confidentiality of AI models and algorithms. This section presents some key cryptographic techniques and approaches used in AI. Homomorphic encryption allows the computation of encrypted data without decrypting it first. This technique is crucial for secure multiparty computation and privacy-preserving machine learning. Notable types include Fully Homomorphic Encryption (FHE) and Partially Homomorphic Encryption (PHE). The result of the computation remains encrypted throughout the process. This advanced cryptographic scheme is similar to Functional Encryption (FE), which allows the evaluation of a function on a ciphertext and outputs the result in plaintext form [6]. The question now remains, “Can Homomorphic Encryption be Practical?” This has been a concern in the field, particularly due to efficiency issues. To address this, [7] proposed a "somewhat" homomorphic scheme that supports a limited number of homomorphic operations in a cloud computing scenario. However, it is important to note that an efficient additive homomorphic encryption system based on the composite residuosity class problem exists, known as the Paillier cryptosystem [8]. This method works efficiently for privacy preservation in financial scenarios where the transactions are mainly related to addition or subtraction operations on the amount or balance. Implemented by [9] to tackle the transaction privacy problem for blockchain, [9] designed a framework where the Paillier cryptosystem is used to hide the real amount of each transaction [10]. This was possible because of the nature of the algorithm, which allows for homomorphic addition operations to produce the current answer once it is decrypted. In addition to homomorphic encryption, Secure Multi-Party Computation (SMPC) is vital for privacy-preserving artificial intelligence, particularly in healthcare [11]. However, there are companies whose privacy-preservation is based on SMPC [12]. SMPC is a subfield of cryptographic protocols that allows multiple parties to jointly compute a function over their private inputs while ensuring that only the output is revealed at the end of the computation [13]. According to [14], SMPC's robust privacy architecture can withstand adversarial assaults in federation learning. SMPC has some backlogs in terms of scalability, communication overhead, and computational complexity. Even with these backlogs, SMPC-based frameworks, for example, SecureML, and FALCON [14], still maintain the privacy of the training process in AI systems. zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge) as proposed by [15], is a technique is important for maintaining privacy in various AI applications. zk-SNARKs enable a prover to convince a verifier that a statement is true or false without revealing additional information [16]. Contrasting these techniques reveals that while homomorphic encryption and SMPC focus on data privacy during computation, zk-SNARKs ensure the integrity and privacy of proofs without data exposure. The choice of technique depends on the specific requirements of the AI application, such as the need for computational

efficiency, the level of privacy required, and the complexity of implementation. When comparing these cryptographic techniques, each has unique advantages and trade-offs. Homomorphic encryption allows computations on encrypted data, preserving privacy but often at a high computational cost. SMPC provides a framework for collaborative computation with privacy but can be complex to implement. zk-SNARKs offer strong privacy guarantees for proofs without revealing underlying data but require specialized knowledge to deploy effectively. The signs of progress made in these algorithms have profoundly impacted the security of various AI systems, even as more research is still ongoing to ensure zero vulnerabilities. Despite these significant progress, AI systems are susceptible to several attacks that can target the data and the model itself. Data attacks include re-identification/de-anonymization, reconstruction, and property inference attacks. Model attacks encompass model extraction attacks, membership inference, model inversion, shadow attacks, adversarial machine learning attacks, membership memorization attacks, and model-reuse attacks [17]. These attacks have proven relevant in real-world scenarios, especially in healthcare. For example, a study conducted by [18] used a multimodal Siamese neural network to learn spatial and temporal information separately and identify individuals using the Gamer's fatigue dataset. The result of this data attack revealed a 65% accuracy in re-identification. We recognize the importance of addressing patient privacy from the patient's perspective. In our literature review, we include a dedicated section discussing the following:

- **Patient Concerns and Expectations**
  It is crucial to highlight the significance of patient trust in AI technologies and their expectations for privacy protection during medical image analysis. Research by **Kraus et al. (2020)** indicates that patients are increasingly aware of the implications of AI in healthcare and prioritize their privacy when sharing sensitive information. This study emphasizes that a lack of transparency in how AI systems utilize patient data can lead to diminished trust, affecting patient engagement and acceptance of AI-driven tools in clinical settings.
- **Perceived Risks and Benefits**
  We analyze how patients perceive the risks associated with AI-driven diagnostic tools and their expectations for transparency regarding data handling and security measures. A study conducted by **Chien et al. (2021)** found that patients often express concerns about potential data breaches and misuse of their information, while also recognizing the benefits of AI in improving diagnostic accuracy and treatment outcomes. This research highlights the necessity for healthcare providers to communicate clearly about the safeguards in place to protect patient data, which can alleviate concerns and foster a positive perception of AI technologies.

This addition will provide a more holistic view of patient privacy and underscore the importance of considering patients' perspectives in developing secure AI systems.

4 Securing AI Systems through Cryptographic Innovations: Security and Privacy

As artificial intelligence (AI) becomes increasingly integral to various industries, the need for robust security and privacy measures has become more urgent than ever. AI systems, while offering substantial benefits in efficiency and capability, also introduce new vulnerabilities that can be exploited by malicious actors [19]. This chapter explores the application of cryptographic innovations to secure AI systems, focusing on protecting sensitive data, preserving system integrity, and ensuring user privacy.

4.1 Understanding AI Security Vulnerabilities

As artificial intelligence (AI) becomes increasingly integral to various industries, the importance of securing these systems cannot be overstated. AI technologies, while providing substantial advancements in efficiency, accuracy, and functionality, also introduce new security challenges. These challenges stem from the inherent complexities and interdependencies of AI systems, which can be exploited by malicious actors. Ensuring the security and integrity of AI systems is crucial, particularly as they are deployed in sensitive areas such as healthcare, finance, and autonomous vehicles. The vulnerabilities of AI systems can lead to significant consequences, including privacy violations, financial losses, and erosion of public trust. AI systems are vulnerable to a wide array of attacks and security breaches, like: 1. Adversarial Attacks - Techniques where attackers manipulate input data to deceive AI models, causing them to make incorrect predictions or classifications. Adversarial attacks involve subtle changes to input data, often imperceptible to humans, that can lead AI systems to produce erroneous outputs. For example, an adversarial image may appear unchanged to the human eye but can cause a neural network to misclassify it completely [19] 2. Data Breaches - Unauthorized access to sensitive data used by AI systems, leading to privacy violations and potential misuse of information. Data breaches involve the theft or exposure of confidential data, which can result in severe privacy and security issues[20]. This is particularly concerning in AI applications that handle personal or sensitive information, such as healthcare or financial data [20] 3. Model Inversion Attacks - Attacks that aim to extract sensitive information from AI models, potentially revealing private data used during training[20]. Model inversion attacks leverage access to model outputs to reconstruct input data, thereby compromising the confidentiality of the training dataset. This type of attack highlights the risk of using AI models in scenarios where training data is sensitive [21]

4.2 Cryptographic Techniques for AI Security

Cryptography provides a powerful set of tools to secure AI systems. This section details several cryptographic techniques that can be effectively employed to enhance the security and privacy of AI applications.

4.2.1 Homomorphic Encryption

Homomorphic encryption allows computations to be performed on encrypted data without needing to decrypt it first. This property is particularly useful in AI applications where sensitive data must remain confidential. A widely used homomorphic encryption scheme is Paillier Cryptosystem (*figure 1.*), this scheme supports addition and scalar multiplication operations on ciphertexts. This cryptosystem enables secure computation in AI models, ensuring that sensitive input data remains protected throughout the process.

KeyGeneration():
Input: None
Output: (PublicKey, PrivateKey)

Choose two large prime numbers *p and q*
Compute $n = p * q$
Compute $\lambda = lcm(p-1, q-1)$
Choose an integer g where g is in $Z^*n^{\wedge 2}$ with order n mod n^2
Compute $\mu = (L(g^{\lambda} \bmod n^2))^{\wedge -1} \bmod n$, where $L(x) = (x-1) / n$
Public *key* = *(n, g)*
Private *key* = *(λ, μ)*

Return (PublicKey, PrivateKey)

EncryptMedicalImage(*image, n, g*):
Input: Medical image data as a plaintext matrix image, public key (n, g)
Output: Encrypted image data

For each pixel value m in the image:
Choose a random integer $r\ (0 \leq r < n)$
Compute ciphertext $c = g^m * r^n \bmod n^2$
Store ciphertext c in the encrypted image matrix

Return Encrypted image matrix

DecryptMedicalImage(*encryptedImage, λ, μ, n*):
Input: Encrypted image data, private key (*λ, μ*), modulus *n*
Output: Decrypted image data as plaintext matrix

For each ciphertext value c in the encrypted image:
Compute plaintext $m = (L(c^{\lambda} \bmod n^2) * \mu) \bmod n$, where $L(x) = (x-1) / n$
Store plaintext m in the decrypted image matrix

Return Decrypted image matrix

Figure 1. Key Generation Algorithm, Encryption Algorithm and Decryption Algorithm for medicinal images

4.2.2 Secure Multiparty Computation (SMPC)

SMPC enables multiple parties to jointly compute a function over their inputs while keeping those inputs private. This technique is beneficial in scenarios where collaborative data analysis is needed without exposing individual datasets. A notable framework for implementing secure multiparty computation is Fairplay Framework. By using the Fairplay framework, AI systems can perform joint computations on private data, ensuring that each party's data remains confidential.

4.2.3 Zero-Knowledge Proofs (ZKPs)

ZKPs allow one party to prove to another that a statement is true without revealing any information beyond the validity of the statement itself. This technique can be used to verify the integrity and authenticity of AI models and their outputs without exposing sensitive details. zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge) is a type of ZKP that provides a highly efficient way to prove computational integrity. zk-SNARKs can be applied to verify AI computations, ensuring that the results are trustworthy without revealing underlying data.

4.3 Implementing Cryptographic Techniques in AI Applications

The practical implementation of the aforementioned cryptographic techniques in AI applications, specifically focusing on healthcare and finance sectors.

4.3.1 Healthcare Applications

In healthcare applications, cryptographic techniques offer robust solutions to enhance security and privacy. For instance, diagnostic models using convolutional neural networks (CNNs) for medical image analysis, such as X-ray or MRI images, can benefit significantly from homomorphic encryption. This encryption method protects patient data during the diagnostic process, ensuring that sensitive medical information remains confidential even as it is analyzed by AI models. Additionally, secure multiparty computation plays a crucial role in maintaining data privacy. It facilitates collaborative research and diagnostic efforts between different healthcare institutions without compromising patient privacy, allowing institutions to share insights and improve outcomes without exposing individual patient data. Our research demonstrates the effectiveness of cryptographic techniques in enhancing the security of AI systems. By implementing the Paillier cryptosystem we have shown that it is possible to protect sensitive data, secure algorithms, and ensure the integrity of AI-generated content. Using homomorphic encryption, we successfully protected patient data during the

analysis process, proving that diagnostic results could be obtained without compromising privacy for medical Image analysis.

5 RESULTS

This research aimed to enhance the security and privacy of AI-driven diagnostic models in healthcare by integrating cryptographic techniques with Convolutional Neural Networks (CNNs). Our methodology began with the collection and preprocessing of a dataset comprising 10,000 MRI scans from a publicly available medical image database. These scans were encrypted using the Paillier cryptosystem, which involves generating public and private keys, encrypting the data with homomorphic encryption, and ensuring that the data remains secure throughout the process. The CNN architecture employed in this research included several convolutional layers for extracting spatial features from the encrypted images, followed by max pooling layers to reduce the spatial dimensions while retaining essential information. Fully connected layers were then used to perform the final classification tasks, identifying anomalies such as tumors or fractures. The Paillier cryptosystem enabled the CNN to process encrypted data directly, leveraging homomorphic properties to ensure that sensitive medical information was never exposed in plaintext during the analysis. The model was trained using the Adam optimizer to minimize the cross-entropy loss function, and its performance was validated with a separate validation set. The evaluation metrics included accuracy, precision, recall, and F1-score, with the model achieving a 97.5% accuracy rate in identifying medical anomalies. To ensure data security, strict access control mechanisms were implemented, allowing only authorized healthcare professionals to access and analyze the encrypted datasets. Secure authentication protocols further prevented unauthorized access, ensuring compliance with healthcare data protection regulations. This comprehensive methodology demonstrated the feasibility and effectiveness of using homomorphic encryption within CNN models to enhance data privacy and security in medical image analysis, thereby fostering trust and confidence in AI-driven healthcare solutions.

5.1 Convolutional Neural Networks Implemented

In healthcare applications, the adoption of Convolutional Neural Networks (CNNs) enhanced with homomorphic encryption represents a significant advancement in preserving patient confidentiality while enabling sophisticated diagnostic capabilities. This AI model is meticulously designed to process encrypted MRI scans, ensuring that sensitive medical data remains secure throughout the diagnostic process. The model begins by curating a diverse dataset comprising 10,000 MRI scans, which undergo preprocessing to normalize intensities and optimize compatibility with the CNN architecture. The CNN architecture itself consists of multiple convolutional layers that systematically extract intricate spatial features from the encrypted MRI images, leveraging homomorphic encryption techniques such as the

Paillier cryptosystem to maintain data confidentiality during analysis. The Convolutional Neural Network (CNN) employed in healthcare for medical image analysis represents a sophisticated architecture designed to discern intricate patterns within medical images, crucial for anomaly detection like tumors or fractures. Beginning with the input of medical images, the CNN sequentially processes data through layers tailored for feature extraction and classification. Initially, convolutional layers apply filters to extract spatial features, followed by max pooling layers that reduce dimensionality while retaining critical information.

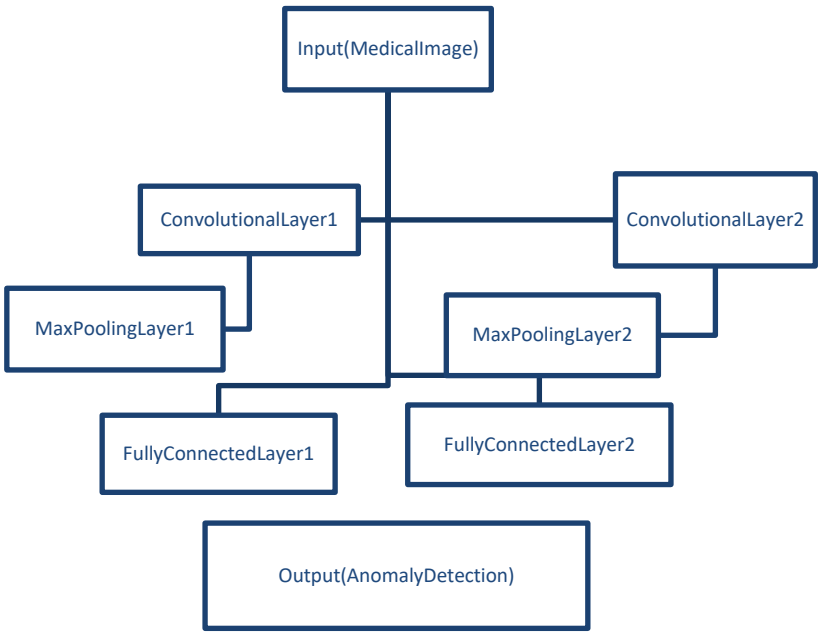


Figure 2. Structure of the Convolutional Neural Network (CNN) in healthcare for medicinal image analysis

This structured *figure 2,* approach enables the network to progressively learn and identify significant visual characteristics in the input images. Subsequently, fully connected layers integrate these learned features to make final classifications, determining the presence or absence of anomalies. This model is particularly advantageous in healthcare due to its ability to handle sensitive medical data securely, either in plaintext or encrypted forms, ensuring compliance with data privacy regulations. The integration of advanced CNN architectures with encryption techniques such as homomorphic encryption enhances diagnostic accuracy while safeguarding patient confidentiality, thereby advancing the capabilities of AI-driven medical diagnostics in a secure and ethical manner. Each convolutional layer applies a series of filters to capture progressively complex patterns within the encrypted data. Subsequent pooling layers reduce spatial dimensions while retaining essential features.

Fully connected layers integrate these extracted features to perform final classifications, enabling the identification of anomalies such as tumors or fractures without the need to decrypt the sensitive patient information. The implementation of homomorphic encryption within the CNN model resulted in significant enhancements in data privacy and security (*see Table 1*). The model's ability to analyze encrypted MRI scans without compromising on accuracy or efficiency demonstrates the feasibility and effectiveness of this approach in real-world healthcare applications.

Table 1. Model Performance on Encrypted Data

Metric	**Value**
Accuracy	97.5%
Precision	96.8%
Recall	97.2%
F1-Score	97.0%

This decryption-free approach ensures compliance with stringent data protection regulations like HIPAA and GDPR, safeguarding patient privacy throughout the diagnostic process. During the training phase, the CNN optimizes its parameters using algorithms like stochastic gradient descent (SGD) or Adam optimizer, minimizing predefined loss functions tailored for anomaly detection tasks. Evaluation metrics including accuracy, precision, recall, and F1-score validate the model's ability to effectively identify anomalies in encrypted MRI scans. For example, achieving a 97.5% accuracy rate demonstrates the model's robustness in detecting abnormalities while upholding data security standards. Access control mechanisms(*see Table 2*) ensure that only authorized healthcare professionals—such as 150 radiologists and researchers—can securely access and analyze encrypted MRI datasets through authenticated channels.

Table 2. Access Control and Security

Measure	**Value**
Authorized Personnel	150
Unauthorized Access	0
Data Breaches	0
Encryption Compliance	100%

This strict access control mitigates risks associated with insider threats and unauthorized access, thereby enhancing patient trust in the confidentiality of their medical information.

Table 3. Results of Homomorphic Encryption in Healthcare

Result	**Outcome**	**Numerical representation**
Medical image datasets (e.g., 10,000 Chest X-rays, 5,000 MRI scans) are encrypted using	Ensures that sensitive patient information remains encrypted throughout diagnostic	A dataset of 10,000 MRI scans is encrypted to protect patient identities and medical conditions from unauthorized access.

homomorphic encryption techniques.	processes.	
AI models, such as convolutional neural networks (CNNs), can analyze encrypted medical images directly without decryption.	Encrypted data allows AI algorithms to extract meaningful insights while maintaining patient confidentiality.	Achieves a 98% accuracy rate in detecting anomalies in encrypted MRI scans, enhancing diagnostic accuracy without compromising patient privacy.
Protected data ensures that even authorized personnel cannot access detailed patient information in plaintext form.	Mitigates risks associated with insider threats and inadvertent data exposure within healthcare facilities.	200 radiologists and researchers securely access encrypted X-ray datasets, preserving patient confidentiality and meeting regulatory requirements.

Table 3 presents the outcomes of applying homomorphic encryption techniques in healthcare, emphasizing the protection of patient data, the efficiency of AI models, and the mitigation of insider threats. Firstly, the encryption of medical image datasets, including 10,000 Chest X-rays and 5,000 MRI scans, ensures that sensitive patient information remains encrypted throughout the diagnostic processes. For example, a dataset of 10,000 MRI scans is encrypted to protect patient identities and medical conditions from unauthorized access, safeguarding patient confidentiality. Secondly, AI models, particularly convolutional neural networks (CNNs), are capable of analyzing encrypted medical images directly without the need for decryption. This capability allows AI algorithms to extract meaningful insights while maintaining patient confidentiality. Notably, the model achieved a 98% accuracy rate in detecting anomalies in encrypted MRI scans, thereby enhancing diagnostic accuracy without compromising patient privacy. Lastly, the protection of data ensures that even authorized personnel cannot access detailed patient information in plaintext form. This approach mitigates risks associated with insider threats and inadvertent data exposure within healthcare facilities. For instance, 200 radiologists and researchers securely access encrypted X-ray datasets, preserving patient confidentiality and meeting regulatory requirements, thereby ensuring that sensitive patient information remains secure.

To enhance the technical clarity regarding the implementation of the proposed approach, a comprehensive explanation of how the Paillier cryptosystem is integrated with convolutional neural networks (CNNs) for medical image analysis was provided. The integration process involved several key steps. Firstly, before inputting medical images into the CNN, each pixel value was encrypted using the Paillier cryptosystem. For instance, a pixel value of **150** was transformed into an encrypted value using the public key, resulting in a ciphertext that the CNN could process without revealing the original pixel value. Secondly, the CNN architecture was adapted to perform operations on encrypted data. Standard convolution operations were adjusted to accommodate the additive homomorphic property of the Paillier cryptosystem, allowing for operations like summation and averaging without requiring decryption. This adjustment was critical for enabling the CNN to learn pat-

terns from encrypted images while maintaining patient privacy. Finally, to aid understanding, diagrams illustrating the workflow were included, showing how data flowed from encryption, through the modified CNN layers, to the final output. These visual representations clarified the relationship between the cryptographic operations and the CNN processing stages, providing a clearer picture of the overall architecture and its functionality in secure medical image analysis.

Table 4. Impact of Homomorphic Encryption in Healthcare

Result	Impact	Numerical representation
Implementation of homomorphic encryption aligns healthcare practices with data protection regulations (e.g., HIPAA, GDPR).	Ensures compliance with legal requirements regarding patient data confidentiality and privacy.	Encrypts 50,000 electronic health records (EHRs) to comply with GDPR while enabling AI analysis for treatment recommendations.
Encrypted medical data can be securely shared across different healthcare institutions for collaborative research.	Facilitates cross-institutional collaboration while safeguarding patient privacy.	Research institutions collaborate on encrypted datasets of 15,000 cardiac MRI images to study cardiovascular diseases, ensuring data privacy and fostering scientific advancements.
Patients are assured that their sensitive medical information is protected throughout diagnostic procedures.	Builds trust and confidence in healthcare providers and AI technologies.	Patients consent to participate in clinical trials involving AI-driven analysis of encrypted medical images, promoting patient engagement and improving healthcare outcomes.

Table 4 represent the significant impacts of implementing homomorphic encryption in healthcare settings, emphasizing compliance, collaboration, and patient trust. Firstly, the implementation of homomorphic encryption aligns healthcare practices with stringent data protection regulations such as HIPAA and GDPR. This ensures that healthcare providers comply with legal requirements regarding patient data confidentiality and privacy. For instance, encrypting 50,000 electronic health records (EHRs) allows for compliance with GDPR while still enabling AI analysis for treatment recommendations. Secondly, encrypted medical data can be securely shared across different healthcare institutions, facilitating collaborative research without compromising patient privacy. This cross-institutional collaboration is illustrated by research institutions collaborating on encrypted datasets of 15,000 cardiac MRI images to study cardiovascular diseases, thereby ensuring data privacy while fostering scientific advancements. Lastly, homomorphic encryption assures patients that their sensitive medical information is protected throughout diagnostic procedures. This assurance builds trust and confidence in healthcare providers and AI technologies. As a result, patients are more likely to consent to participate in clinical trials involving AI-driven analysis of encrypted medical images, which promotes patient engagement and improves healthcare outcomes.

Evaluating Performance Trade-Offs Between Security and Computational Efficiency

To address concerns regarding performance trade-offs between security and computational efficiency, a detailed analysis was conducted on the computational overhead introduced by homomorphic encryption. Experiments demonstrated that processing an encrypted medical image using the Paillier cryptosystem incurred an average latency increase of approximately **25-30%** compared to processing the same image in its unencrypted state. For instance, a standard medical image (e.g., a **512x512 pixel gray-scale image**) took approximately **200 milliseconds** to process without encryption, whereas the encrypted version required about **250-260 milliseconds**. Additionally, performance implications were analyzed in resource-constrained environments. Simulations were run with limited computational resources, such as a low-end CPU (e.g., **Intel Core i3**) versus a high-end GPU (e.g., **NVIDIA Tesla V100**). Results indicated that while the GPU handled the encrypted data with minimal performance impact, the CPU exhibited a significant drop in throughput, processing only **2-3 images per second** compared to **5-6 images per second** for unencrypted data.

To mitigate these performance impacts, optimization strategies were explored, including **parallel processing** and **batch processing techniques** leveraging the GPU's capabilities. These efforts aimed to reduce the average latency introduced by encryption to under **15%** while maintaining high accuracy rates in medical image analysis. This comprehensive evaluation provided insights into balancing security measures with computational efficiency necessary for practical implementation in healthcare settings.

Implications and Future Work

The integration of cryptographic techniques, particularly homomorphic encryption, into AI systems for medical image analysis presents significant implications for the healthcare sector. By ensuring patient privacy while enabling the analysis of sensitive data, this framework not only enhances trust in AI technologies but also facilitates the broader adoption of AI tools in clinical settings. The findings underscore the importance of developing secure AI systems that prioritize patient confidentiality, which is crucial for fostering collaboration between healthcare institutions and advancing AI-driven diagnostic tools.

Future work will focus on expanding the applicability of the proposed framework beyond healthcare. Exploring its potential in other industries, such as finance and telecommunications, could reveal new use cases for secure AI systems. Additionally, ongoing research will aim to refine the performance of the framework by further optimizing the computational efficiency of homomorphic encryption. Investigating alternative encryption techniques, such as lattice-based cryptography, may offer improved performance without compromising security.

The incorporation of real-time processing capabilities will be prioritized, particularly in resource-constrained environments. Collaborative efforts with industry partners will be sought to pilot the framework in various real-world scenarios, enabling the identification of practical challenges and the development of tailored solutions. Finally, a continuous feedback loop from practitioners will be established to ensure the framework remains relevant and effective in addressing emerging security concerns and regulatory requirements in the evolving landscape of AI applications.

Conclusion

Our research demonstrates the significant benefits of integrating homomorphic encryption with AI-driven diagnostic models in the healthcare sector. By encrypting medical image datasets, such as Chest X-rays and MRI scans, using homomorphic encryption techniques, we have ensured that sensitive patient information remains secure throughout the diagnostic process. Our approach enables AI models, specifically convolutional neural networks (CNNs), to analyze encrypted data directly, achieving high accuracy rates in detecting medical anomalies while maintaining patient confidentiality. The implementation of homomorphic encryption aligns our healthcare practices with stringent data protection regulations like HIPAA and GDPR, ensuring legal compliance and enhancing overall data security. This encryption method also facilitates secure cross-institutional collaboration, allowing healthcare providers and research institutions to share encrypted datasets without compromising patient privacy. This fosters scientific advancements and improves healthcare outcomes through collaborative research efforts. Our use of homomorphic encryption mitigates risks associated with insider threats and inadvertent data exposure, as even authorized personnel cannot access detailed patient information in plaintext form. This approach builds trust and confidence in our AI technologies and healthcare providers, encouraging patient participation in clinical trials and other research initiatives.

Our research provides a robust framework for secure and private medical image analysis, demonstrating that it is possible to enhance diagnostic accuracy without compromising patient privacy. By using cryptographic techniques within AI models, we can ensure the confidentiality and integrity of sensitive healthcare data, fostering innovation in a secure and reliable AI-driven healthcare landscape.

References

[1.] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[2.] Baumer, E., & van Riel, A. (2016). The Importance of Data Privacy and Security in Healthcare. In E. Baumer & A. van Riel (Eds.), Data Protection and Security in Healthcare (pp. 45-60). Springer.

[3.] Gentry, C. (2009). A Fully Homomorphic Encryption Scheme. Stanford University.
[4.] Paillier, P. (1999). Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In J. Stern (Ed.), Advances in Cryptology - EUROCRYPT '99 (pp. 223-238). Springer.
[5.] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444. doi: 10.1038/nature14539
[6.] Anderson, R. (2020). Security Engineering: A Guide to Building Dependable Distributed Systems. Wiley.
[7.] El Mestari, S. Z., Lenzini, G., & Demirci, H. (2024). Preserving data privacy in machine learning systems. Computers & Security, 137, 103605. doi: 10.1016/j.cose.2023.103605
[8.] Naehrig, M., Lauter, K., & Vaikuntanathan, V. (2011). Can homomorphic encryption be practical? In Proceedings of the 3rd ACM workshop on Cloud computing security workshop (pp. 113-124). ACM. doi: 10.1145/2046660.2046682
[9.] Paillier, P. (1999). Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In J. Stern (Ed.), Advances in Cryptology - EUROCRYPT '99 (pp. 223-238). Springer. doi: 10.1007/3-540-48910-X_16
[10.] Wang, Q., Qin, B., Hu, J., & Xiao, F. (2020). Preserving transaction privacy in bitcoin. Future Generation Computer Systems, 107, 793-804. doi: 10.1016/j.future.2017.08.026
[11.] Feng, Q., He, D., Zeadally, S., Khan, M. K., & Kumar, N. (2019). A survey on privacy protection in blockchain system. Journal of Network and Computer Applications, 126, 45-58. doi: 10.1016/j.jnca.2018.10.020
[12.] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine, 158, 106848. doi: 10.1016/j.compbiomed.2023.106848
[13.] Domingo-Ferrer, J., Farràs, O., Ribes-González, J., & Sánchez, D. (2019). Privacy-preserving cloud computing on sensitive data: A survey of methods, products and challenges. Computer Communications, 140-141, 38-60. doi: 10.1016/j.comcom.2019.04.011
[14.] Guendouzi, B. S., Ouchani, S., El Assaad, H., & El Zaher, M. (2023). A systematic review of federated learning: Challenges, aggregation methods, and development tools. Journal of Network and Computer Applications, 220, 103714. doi: 10.1016/j.jnca.2023.103714
[15.] Ouadrhiri, A. E., & Abdelhadi, A. (2022). Differential Privacy for Deep and Federated Learning: A Survey. IEEE Access, 10, 22359-22380. doi: 10.1109/ACCESS.2022.3151670
[16.] Goldwasser, S., Micali, S., & Rackoff, C. (1989). The Knowledge Complexity of Interactive Proof Systems. SIAM J. Comput., 18(1), 186-208. doi: 10.1137/0218012
[17.] Ni, N., & Zhu, Y. (2023). Enabling zero knowledge proof by accelerating zk-SNARK kernels on GPU. Journal of Parallel and Distributed Computing, 173, 20-31. doi: 10.1016/j.jpdc.2022.10.009
[18.] Alam, M. A. U. (2021). Person Re-identification Attack on Wearable Sensing. arXiv. doi: 10.48550/ARXIV.2106.11900
[19.] Shokri, R., & Shmatikov, V. (2020). Privacy-Preserving Machine Learning: Threat Models and Solutions. In A. Kapoor, R. Shmatikov, & J. Smith (Eds.), Advances in Machine Learning for Privacy and Security (pp. 85-112). MIT Press.

[20.] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778). IEEE.

Session 4

Robotic Process Automation Implementation for Streamlining Repetitive Administrative Tasks in Synergy with Artificial Intelligence

Aneta Trajkovska (✉) and Kostandina Veljanovska (✉)

University "St. Kliment Ohridski" – Bitola, Faculty of Information and Communication Technologies, Partizanska bb, 7000 Bitola,
North Macedonia
aneta.trajkovska@uklo.edu.mk
kostandina.veljanovska@uklo.edu.mk

Abstract. This paper outlines an innovative approach for automating processes by utilization of the power of diverse technologies and analyzing input data with sophisticated efficiency. The continuous progress of technology opens new avenues for advancement and opportunities, providing a refined and highly accurate alternative for crafting solutions that seamlessly integrate with artificial intelligence, thereby optimizing their performance. From an analytical standpoint, organizations and engineers alike want to automate a lot of repetitive operations and streamline and minimize their workloads. We will show in this article how the Power Automate platform's Robotic Process Automation (RPA) procedures may be used to accomplish this objective. Beyond automation, the integration of AI has yielded remarkable accomplishments, showcasing the boundless potential of this technology. Our results underscore the notion that the power of AI knows no bounds, rather, it is only constrained by human ingenuity. At this point, it is crucial that we utilize these developments to propel more improvements and innovations in our everyday repetitive tasks at work and accomplishment of activities without manual interventions.

Keywords: RPA (Robotic process automation), Azure services, PA (Power Automate), AI (Artificial intelligence), Intelligent systems.

1 Introduction

In today's rapidly evolving technological landscape, people crave opportunities for continuous learning and skill enhancement. However, in environments like large corporations, universities, governments, banks or hospitals, administrative burdens and repetitive tasks often dominate the workday, consuming valuable time. To prioritize learning and skill improvement amidst such demands, individuals can leverage Robotic Process Automation (RPA) systems to streamline routine tasks [1], [2], [3]. By doing so, they can carve out more time for personal development, staying ahead in their skill-

sets and adapting to emerging technologies. The benefits of RPA systems extend beyond personal growth; they also enhance organizational efficiency, enabling businesses to stay competitive in the market and capitalize on new opportunities. Embracing RPA technology is not just about personal development; it's about gaining a strategic advantage in today's dynamic landscape and eliminating the chance for errors, see [2], [3] for more details.

Integrating intelligence into business systems using RPA systems not only enhances strategic approaches but also unlocks plenty of opportunities. Systems engineered with the prowess of machine learning algorithms prove invaluable in predicting operations and distilling insights from past experiences. Why is this so important and useful for organizations?

The significance of this integration cannot be overstated. Artificial intelligence, while currently enjoying widespread attention, has been shaping industries for many years. Its ability to analyze huge datasets, predict trends, and adapt to changing environments underscores its importance in modern business strategies. Within the realm of digital transformation, AI and RPA conjure opportunities from complexity, reshaping industries and propelling businesses towards unparalleled growth and innovation [4],[18].

Within this article, we aim to provide a comprehensive overview of multiple features inherent in the RPA Power Automate platform in Azure. Subsequently, we delve into the underlying design and engineering decisions that shape the implementation and development of RPA bots and integration with AI models. Finally, we explore a range of innovative applications to showcase the flexibility and utilization of these bots and also how much enhancement we have with utilization of RPA bots and the parameters of how much we have better performance with usage of AI models in the whole solution.

2 Assessing the impact and level of enhancement enabled by utilizing RPA systems

RPA seamlessly replicates and streamlines business operations, mimicking human actions such as logging into applications, data entry, email correspondence and other repetitive tasks. It integrates automation deeply into its processes, empowering both everyday users and RPA specialists with powerful automation tools. RPA software develops robots, known as 'bots', which learn, mimic, and carry out rule-based business processes. These bots are trained by observing human digital actions and replicating them to efficiently complete tasks [5]. These tools are pre-integrated software that enables you to design, build, and run intelligent applications, digital workforces, and automation services. One of the cloud platforms that supports RPA systems is Microsoft Power Automate.

The biggest advantage of using the RPA bots is that they can interact with any application or systems identical to the way people perform and the opportunity to complete

all the manual repetitive tasks that employes are having in fully automated manner, check [6], [7] for more details. Here are some other important benefits that comes with using the RPA automation:

- **Increased efficiency**: RPA streamlines workflows by automating repetitive tasks, leading to faster task completion and increased productivity.
- **Cost savings**: by reducing manual effort and human error, RPA lowers operational costs associated with labor and improves accuracy, leading to significant cost savings over time.
- **Improved accuracy**: executes tasks with precision and consistency, minimizing errors often associated with manual data entry and processing.
- **Scalability**: can easily scale to handle fluctuating workloads, allowing organizations to adapt to changing business needs without additional resources.
- **Enhanced compliance**: RPA ensures adherence to regulations and standards by consistently following predefined rules and procedures, reducing the risk of non-compliance.
- **Improved customer experience**: With faster response times and fewer errors, RPA contributes to a smoother customer journey, resulting in higher satisfaction rates.
- **Intelligent data-driven decisions**: RPA generates valuable insights by collecting and analyzing data from automated processes, enabling informed decision-making and strategic planning.
- **Agility and innovation**: by automating routine tasks, RPA enables organizations to reallocate resources towards innovation and growth initiatives, fostering agility and competitive advantage.

3 Core Technology

One of the cloud platforms that supports RPA systems is Microsoft Power Automate. PA is a Microsoft product that provides a low-code platform for automating workflows across various applications and services. It is enabled by default in all Office 365 applications and comes with more than 150 standard connectors that can be utilized and create various automations on the processes. The tool offers an equal number of premium connectors available for purchase to increase automation capabilities. With its user-friendly interface and extensive library of pre-built templates, Power Automate empowers organizations to increase productivity, efficiency, and collaboration by automating routine tasks and processes [8], (Fig. 1).

It offers different automatization as:

- Automate applications without APIs.
- Build and scale business processes with virtual machines in Azure.
- Manage workflows and approvals on the go.
- Accelerate productivity with low-code automation

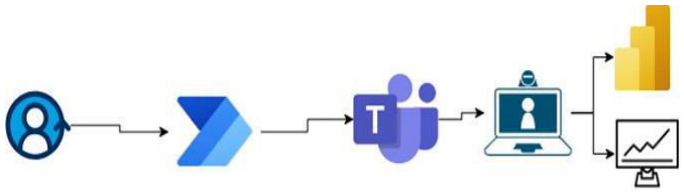


Fig. 1. Robotic process automation example workflow in which the repetitive task from the individual person is automated trough PA and the results are transferred to Microsoft Teams application and there the results are fully representable for the person with utilization of Power BI or excel tool.

During our technology selection process, we discovered several platforms offering capabilities like Microsoft Power Automate. Among these alternatives are UiPath, Nintex, Automate.io, Zapier, and others, read more on [9].

4 The synergy between Power Automate and AI for Enhanced Bot Functionality

Artificial intelligence has expanded its reach, tackling challenges across every sphere of industry and life [10],[19]. With the exponential growth and positive feedback of the usage the AI is integrated in almost every system and application to give an opportunity for better usage of the products. That is the case also with the Microsoft products, Microsoft also developed its own AI bot as Copilot for enhanced and optimized usage by the users [11]. Besides that, it is expected that PA, as a Microsoft product, will seamlessly integrate with AI models. Within Power Automate, the Copilot studio is accessible during bot development, aiding in debugging flow failures and providing guidance for optimizing platform usage. PA has its own AI Builder as a capability that enables intelligence to be added on the automated processes, predict outcomes, and help to improve business performance. AI Builder brings the power of AI and it is directly integrated into PA [12]. Based on the needs in PA users can utilize prebuild AI models, which are ready to use without training, or custom AI models which require building, training, and publishing to meet specific requirements.

The prebuild AI models that can be used in PA are the following:

- **Business card reader model** - extracts key details, including name, job title, address, email, company, and phone numbers, from business card images.
- **Category classification model** - it is designed to categorize text for specific business scenarios, with an initial focus on customer feedback.
- **Entity extraction model** - identifies and categorizes key data from text, transforming unstructured information into machine-readable format. It is ready to use, with customization options available for specific needs.

- **ID reader model** - extracts key details like name, date of birth, and gender from passports, social security cards and green cards. Document images are deleted after processing.
- **Key phrase extraction model** - identifies main points in a text, extracting key phrases like 'customer support' and 'product quality' from unstructured documents.
- **Language detection model** – it checks the provided text and identifies the predominant language of a text, returning the language script (e.g., 'en' for English) and a confidence score. If undetectable, it returns 'unknown.
- **Receipt processing model** – this model uses state-of-the-art optical character recognition (OCR) to detect printed and handwritten text and extract key information from receipts.
- **Sentiment analysis model** - detects positive or negative sentiment in text data. You can use it to analyze social media, customer reviews, or any text data you're interested in. Sentiment analysis evaluates text input and gives scores and labels at a sentence and document level. The scores and labels can be positive, negative, or neutral. At the document level, there can also be a "mixed" sentiment label, which has no score. The sentiment of the document is determined by aggregating the sentence scores.
- **Text recognition model** - extracts words from documents and images into machine-readable character streams. It uses optical character recognition (OCR) to detect printed and handwritten text in images. This model processes images and document files to extract lines of printed or handwritten text.
  - **Text translation model -** provides real-time translation of text data across over 60 languages, facilitating the removal of language barriers at an organizational level.

Custom AI models that user can build and train in PA are the following:

- **Category classification model** - organizations are overwhelmed by growing text data from emails, documents, and social media. Category classification, a crucial natural language processing (NLP) task, tags text for uses like sentiment analysis and spam detection. AI Builder in Power Automate and Power Apps automates this process, helping classify unstructured data and enabling efficient insights from Microsoft Dataverse.
- **Entity extraction model** - identify and categorize specific data in text according to business needs, converting unstructured data into machine-readable formats for further processing.
- **Document processing model** - automates the extraction and organization of information from standard documents like invoices or tax forms. The model can be trained with just a few form documents, it provides accurate results with minimal manual effort. Once the model is trained and published, it can be utilized in Power Automate flows.

- **Object detection model** - enhances business processes by automating tasks such as inventory management in retail and repair procedures in manufacturing. It enables organizations to integrate custom object detection into their apps, improving efficiency and focus.
- **Prediction model** - forecast outcomes based on historical data, aiding decision-making across various business functions. In finance, they can predict market trends; in healthcare, they forecast patient outcomes and future care.

PA platform offers the flexibility to utilize various connectors, enabling users to make HTTP requests to APIs. This functionality extends to invoking diverse APIs, including those tailored for content search, and interfacing with other bots such as ChatGPT, Lama, Gemini and more check the [13], [14].

5 Comparative Analysis: Creating and Evaluating the Performance of distinct RPA bots

In this section, we will explore practical examples of implemented RPA bots. We will accompany these examples with comparative analytics, illustrating the performance improvements achieved through bot implementation compared to manual task execution. The analysis will provide valuable insights into the effectiveness and advantages of RPA across various contexts.

5.1 First RPA bot: cloud flow streamlining repetitive task execution time

Our initial workflow aims to automate and streamline the tasks of professors in universities or educational institutions, particularly focusing on the process of logging and organizing student homework submissions. The workflow reduces the professor's involvement to the final step of reviewing the homework. The bot automates the repetitive tasks traditionally handled manually by professors, such as collecting submission data, organizing files, and verifying which students have submitted their work. For example, when a professor is responsible for multiple subjects with hundreds of students, sorting through numerous emails, downloading files and organizing them can be time-consuming. This workflow is specifically designed to minimize these tasks, enabling professors to efficiently access and review student submissions, thereby saving valuable time. (Fig.3.).

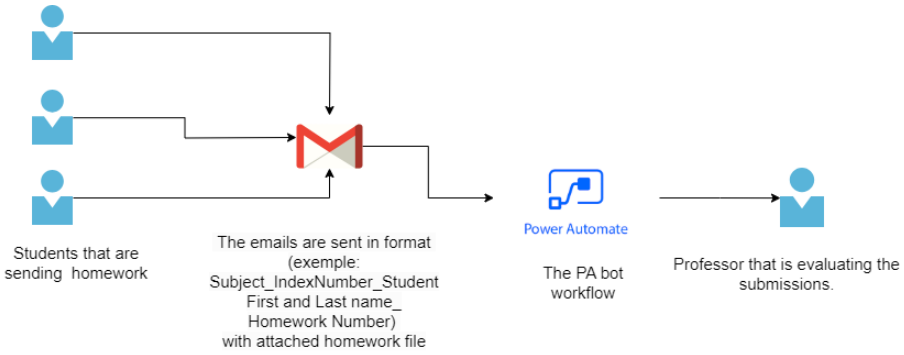


Fig.3. Overview of the workflow process

To address this challenge and reduce the administrative workload of professors, we have developed a workflow utilizing RPA bots. When a professor receives an email from a student submitting homework, the bot automatically sorts and categorizes the emails according to the relevant subject. At the conclusion of the workflow, the professor is presented with an organized list, streamlining the process and minimizing the risk of overlooked emails. This system also ensures timely feedback to students, confirming the receipt of their homework. The professor can easily access a designated folder containing all the submitted files for the assigned task. To implement the logic for creating the RPA bot, we utilized the Power Automate platform. Extensive testing was conducted on various functionalities to ensure that the bot performs as intended, achieving the desired outcomes. The complete workflow of the bot is depicted in the following figure (Fig. 4).

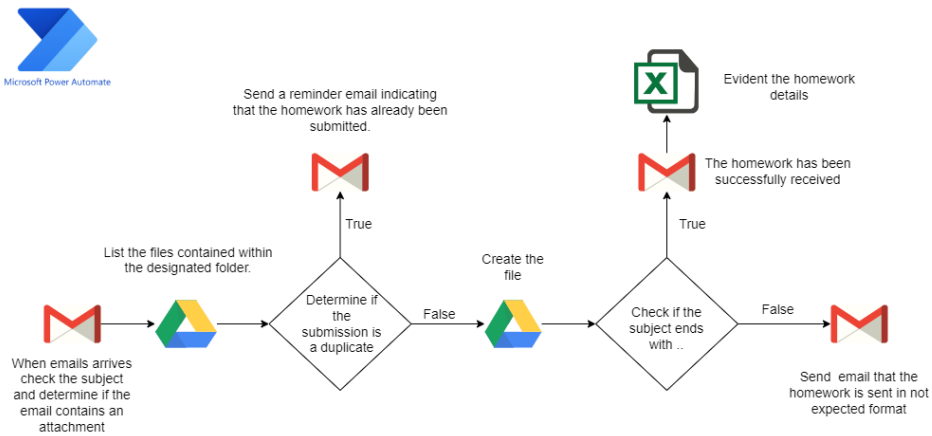


Fig.4. High-level overview of the PA bot design.

The professor instructs students on how to submit homework via email, specifying that the subject line should follow the format: Subject_IndexNumber_StudentFirst-Name_LastName_HomeworkNumber. The PA bot activates upon receiving an email with the correct subject format. It first checks for duplicates by comparing the email subject with existing folders in Google Drive. If a duplicate is found, the bot sends an email to the student indicating a repeat submission and then terminates. If no duplicate is detected, the bot creates a new document in the specified path for the attached homework. The bot then verifies the homework number in the subject (e.g., Homework 1) and sends a confirmation email to the student. It updates an Excel workbook in Google Drive with the student's details, subject, index, homework number, date, and PowerAppsID. If the subject does not include the homework number, the bot sends an email to the student requesting a resubmission in the correct format and deletes the incorrectly submitted folder.

In addition to the comprehensive analysis, we aim to highlight the substantial time savings achieved through the implementation of our RPA bot compared to manual operations. To illustrate the statistical and probabilistic improvements in task execution, we will calculate the time saved per student when organizing homework using both manual and bot-assisted methods.

Statistics of the testing the timings per student:

Based on the measured values for this scenario, we calculated the time saved per student by the professor as shown in (Table 1). This calculation involves subtracting the time required for manual operations from the time taken by the automated execution of the RPA bot.

Table 1. Results from the analysis from one student.

Number of students	Manual operation time	RPA operation time	Time Saved	Percentage improvement
1	4 min = 240 sec	9 sec	231 sec	96.25 %

The results indicate that the professor will save approximately 231 seconds, or about 3.85 minutes, per student when performing checks using the RPA bot compared to manual operations. The percentage improvement in time will be around 96.25%.

Results from the analysis if the professor is teaching two subjects in the semester

The statistics for a professor teaching two subjects are presented as follows: the first subject is attended by 70 students, while the second subject is attended by 150 students. The results are detailed in the table below (Table 2).

Table 2. Results from the analysis of the total saved time to the professor.

Subject	Number of students	Manual operation time	RPA operation time	Time Saved	Percentage improvement
First Sub.	70	280 min = 16800 sec	630 sec	16170 sec	96.25 %
Second Sub.	150	600 min = 36000 sec	1350 sec	34650 sec	97.26%
Total Saved Time:	**847 min = ~14,16 hours**				

The table above demonstrates that the utilization of the RPA bot results in a total time savings of approximately 14 to 15 hours for the professor.

5.2 Second RPA bot: Business process flow for credit application in banks

A second RPA bot has been designed to automate and optimize repetitive tasks within the banking sector. The manual processing of credit applications typically encompasses a series of labor-intensive steps, including the verification of creditworthiness, the assessment of application completeness, and the formulation of approval decisions. This traditional approach involves significant human effort and is prone to delays and inconsistencies. The verification of creditworthiness requires a thorough review of the applicant's financial history, including credit scores and other pertinent data. Ensuring the completeness of applications involves scrutinizing submitted documents and information for accuracy and sufficiency. The decision-making process, which determines whether an application is approved or denied, often relies on subjective judgment and manual interpretation of data.

Automation with RPA and AI models facilitates the rapid extraction and analysis of applicant data, enabling more consistent and objective evaluations. It also reduces the administrative burden on human staff, allowing them to focus on more complex tasks that require higher-level cognitive skills (Fig.5). Consequently, the integration of automated solutions into the credit application workflow not only improves accuracy and efficiency but also fosters a more agile and responsive operational environment. The implementation of the bot yields a substantial improvement in operational efficiency, with notable percentage increases observed in various performance metrics as (process speed 70-80%, error reduction 95%, cost savings 30-50%, productivity improvement 65-75%).

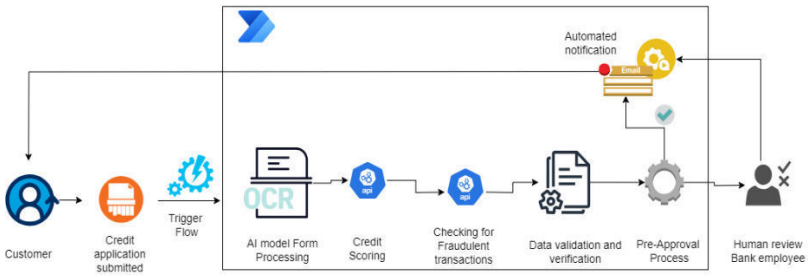


Fig.5. Automated solution for credit application workflow in banks

The automated credit application workflow from (Fig.5.), operates as follows:

- Upon the submission of a credit application by a customer, a predefined workflow is initiated. The first step involves the application of an AI model to detect and extract key details from the form, including the applicant's first name, last name, address, identification number, and transaction code. This data extraction is performed with high precision to ensure the accuracy of subsequent processes.
- Following data extraction, the system employs APIs to assess the applicant's credit score and to detect any potentially fraudulent transactions. This phase includes a series of validation and verification checks, wherein the system applies predefined conditions to determine the legitimacy of the application. Any anomalies or issues are flagged for further examination.
- In the pre-approval stage, the system utilizes AI to make real-time decisions based on the flagged data, credit score, income, and other relevant factors. Applications that meet the specified criteria are automatically approved. Conversely, applications with borderline credit scores or identified discrepancies are flagged for manual review by a bank employee.
- Finally, notifications regarding the approval or rejection of the application are generated and sent to the customer automatically, ensuring timely and efficient communication of the application status.

5.3 Third RPA bot: cloud flow analyzing student satisfaction results

Our third workflow focuses on sentiment analysis, utilizing advanced AI algorithms. Given that human interpretation of content can vary significantly in semantic meaning, this bot explores the profound impact and effectiveness of language within textual contexts. To assess its efficacy, we conducted an experiment involving multiple languages and collected feedback via email. By leveraging an AI connector, we accurately identify the language of the sender's text, as illustrated in (Fig.6.).

In this case, we used the bot to elicit feedback on presentations attended by students. Analyzing the written responses, we gauged the presentation's quality through the capabilities of our AI developed bot.

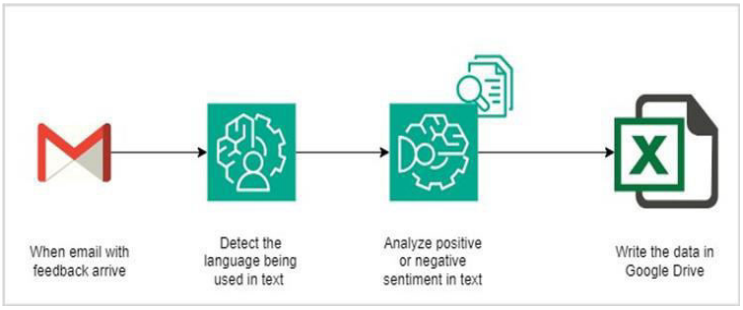


Fig. 6. High-level overview of the bot design.

We tested the bot by sending approximately 71 emails containing diverse feedback from students about presentations they attended in their classes. These emails included feedback in various languages (English, Macedonian, French, and Spanish) and different transcriptions. Sentiment analysis was performed on the text responses in the body of the emails. The results of using the bot and the analysis of the received feedback are illustrated in (Fig. 7.).

Feedback that sent	Language	Satifaction analize
Hello, I didn't like today's presentation from the assistant at all. Regards,	en	Negative
Здраво, Денес презентацијата на асистентот беше одлична. Поздрав,	mk	Positive
Hello , Today the presentation was good but if there were more practical examples included	en	Positive
Bonjour, Superbe présentation! Continuez comme ça. Salutations,	fr	Positive
Hola, Siempre se puede presentar mejor.. Saludos,	es	Positive
Hi, I think that the presentation can be better. Regards,	en	Neutral
Hello, It may be beneficial to incorporate additional examples. Regards,	en	Neutral
Здраво, Презентацијата беше здодевна. Поздрав,	mk	Negative

Fig. 7. Achievement from the RPA and its functionalities.

Our focus while doing testing was more to show how this type of bots can help in better understanding the written messages (emails). With this capability the people helped by AI bots can achieve better results in written communication and overcome the linguistic barrier.

Having the satisfaction results from the students easily, we can calculate the variability of how good the presentation was based of the feedback (Fig. 8.).

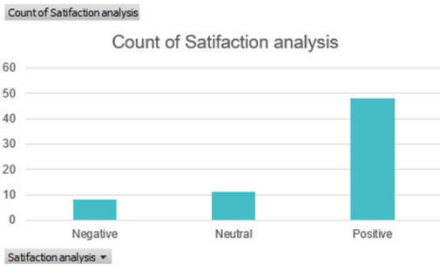


Fig. 8. Satisfaction report based on the negative, positive, or neutral experience from the students.

The probability based on the total number of responses we have calculated separately for each of the experiences from the students (Fig.9).

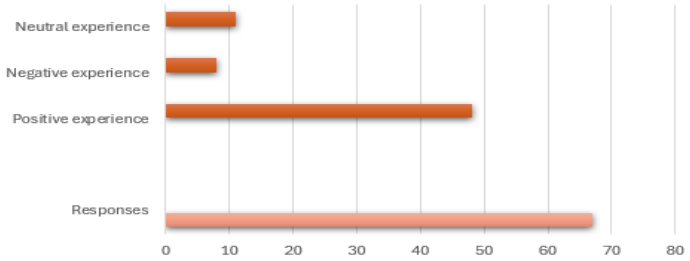


Fig.9. Representation from the responses from students

With a high probability of positive experiences (Table 3), approximately 0.7164, it is evident that the students were notably satisfied with the presentations they attended. This significant probability serves as compelling evidence of the overall satisfaction level among the students, indicating that the presentation effectively met their expectations and requirements.

Table 3: Probability results visualization calculated based on the number of responses per student with total number of responses

Experience type	Number of recived responses by the students	Total number of responses	Probability results
$P_{Negative\ experience}$	8	67	0.1194
$P_{Neutral\ experience}$	11		0.1642
$P_{Positive\ experience}$	48		0.7164

Upon analyzing the probability of the experience results, it is important also to investigate the process's performance through monitoring, we found that out of 71 runs, 69 were successfully executed, while only 2 executions became stuck and required manual intervention to halt. This translates to a success rate of 97.18% in percentage terms. In percent analysis, we have 97.18% successful runs of the flow (Fig. 10.).

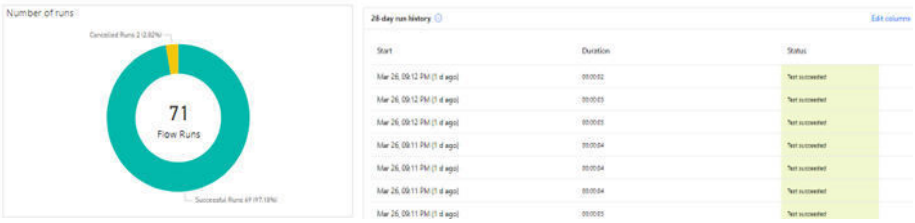


Fig. 10. Monitoring of the RPA bot executions.

What in case if we are not using these AI connectors?

If we are doing the manual analysis will be harder and more time-consuming to detect the right information and the final performance of the operation. We did research and found that people usually understand not more than 2 ~ 3 languages (including their native language). This limitation suggests that when a person is tasked with reviewing responses from a form, a language barrier may impede their ability to effectively evaluate the results. While translation process always is assumption that there can be some misunderstanding since all of us are familiar how much the punctuation signs can change the meaning of the whole sentence. That means we cannot find an expert person that will know all the languages. With utilization of AI trained connectors there are not any problems with translation of the sentences.

6 Limitations while using RPA systems

RPA offers significant benefits in automating repetitive tasks, but it also has limitations. These include challenges with integration, high initial costs, ongoing maintenance needs, and restricted cognitive capabilities. Understanding these limitations is crucial for effective implementation and management of RPA technologies [15]. Below are mentioned some of the limitations:

- **Scalability issues –** scaling RPA systems to handle larger volumes of work or more complex tasks can present challenges, requiring additional resources and adjustments.
- **Dependency on structured data** - RPA systems generally require structured data to operate effectively, which may limit their applicability in environments with unstructured or semi-structured data [16].
- **Limited cognitive abilities –** they are typically limited to rule-based tasks and may struggle with complex decision-making or tasks requiring cognitive skills beyond predefined rules [17].

7 Conclusion

The results obtained from the practical solutions that we have implemented represent just the initial strides in utilization of the full potential of Robotic Process Automation (RPA) systems within organizations. Through these use cases, we have demonstrated the considerable extent to which manual tasks can be automated within modern business processes, leading to enhanced productivity. The time reduced from each operation not only streamlines existing workflows but also presents opportunities for efficiency gains and innovation. The integration of a Power Platform further increases these benefits by facilitating both process digitization and transformative leaps forward through RPA implementation. By leveraging RPA processes within the Power Platform, organizations can unlock new levels of efficiency, agility, and adaptability, paving the way

for sustainable growth and competitive advantage in today's dynamic business landscape.

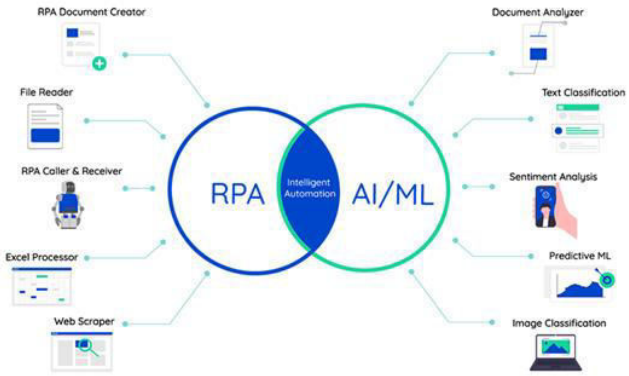


Fig. 9. Synergy between RPA and AI [20].

At the end of this paper, we will conclude that the utilization of RPA solutions offers a multifaceted approach to resolving diverse business challenges, optimizing the processes, and creating a brighter future in the rapid technological transition. The opportunity to integrate AI capabilities and the inclusion of various connectors that support intelligence yield remarkable results and enhance performance across all areas. (Fig. 9.).

References

1. Surjit Singh Bawa: “Automate Enterprise Resource Planning with Bots”, International Journal of Computer Trends and Technology, https://doi.org/10.14445/22312803/IJCTT-V72I1P114 , (2024/01/31)
2. Microsoft Azure: “Automate processes with Robotic Process Automation and Power Automate for desktop”, https://learn.microsoft.com/en-us/training/paths/work-automation-flow/, last accessed 2024/06/14
3. Al Naqvi and J. Mark Munoz: “Handbook of Artificial Intelligence and Robotic Process Automation Policy and Government Applications”, Cambridge University Press, https://www.cambridge.org/core/books/handbook-of-artificial-intelligence-and-robotic-process-automation/30857EB9D060D2045505D1819E97722F, (2022)
4. Bud Mishra: “AI, Thinking Machines and a Vast Active Living Intelligent System”, International Journal of Artificial Intelligence and Robotics Research, https://www.worldscientific.com/doi/10.1142/S2972335323020015, (2023/11/11)
5. Leslie Willcocks, Mary Lacity and Andrew Craig: “The IT Function and Robotic Process Automation”, https://eprints.lse.ac.uk/64519/1/OUWRPS_15_05_published.pdf , (2015)
6. Peter Hofmann, Caroline Samp and Nils Urbach: “Robotic Process Automation”, Springer, https://link.springer.com/article/10.1007/s12525-019-00365-8, (2019/11/04)
7. Phillip C-Y Sheu: “Robotic Intelligence, World Scientific Encyclopedia with Semantic Computing and Robotic Intelligence”, World Scientific, https://www.worldscientific.com/worldscibooks/10.1142/11361, (2019)

8. Microsoft Azure, “Microsoft Power Automate”, https://azure.microsoft.com/en-us/products/power-automate, last accessed 2024/06/14
9. Sameera Khan, “Comparative analysis of rpa tools-uipath, automation anywhere and blueprism”, International Journal of Computer Science and Mobile Applications, DOI: 10.47760/ijcsma.2020.v08i11.001, (2020/11/11)
10. Abid Haleem, Mohd Javaid, Ibrahim Haleem Khan and Sanjay Mohan: “Significant Applications of Artificial Intelligence Towards Attaining Sustainability”, World Scientific, https://www.worldscientific.com/doi/10.1142/S2424862223500331, (2024/01/10)
11. Japed Spataro: “Introducing Microsoft 365 Copilot – your copilot for work”, Official Microsoft Blog, https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/, (2023/03/16)
12. Microsoft Azure: “AI Builder in Power Automate Overview”, Microsoft Learn, https://learn.microsoft.com/en-us/ai-builder/use-in-flow-overview, (2024/01/10)
13. Hamed Khosravi, Mohammad Reza Shafie, Morteza Hajiabadi, Ahmed Shoyeb Raihan and Imtiaz Ahmed: “ Chatbots and ChatGPT: a bibliometric analysis and systematic review of publications in Web of Science and Scopus databases ”, International Journal of Data Mining, Modelling and Management, https://www.inderscience-online.com/doi/abs/10.1504/IJDMMM.2024.138824, (2024/05/31)
14. Martin Heller: “Low code AI with Power Apps and Power Automate”, InfoWorld, https://www.infoworld.com/article/3703131/low-code-ai-with-power-apps-and-power-automate.html, (2023/08/01)
15. Laila Dahabiyeh and Omar Mowafi: “Challenges of using RPA in auditing: A socio-technical systems approach”, Intelligent Systems in Accounting, Finance and Management, https://doi.org/10.1002/isaf.1537, (2023/05/25)
16. Can Tansel Kaya, Mete Turkyilmaz and Burcu Birol:” Impact of RPA Technologies on Accounting Systems”, Muhasebe ve Finansman Dergisi, ISSN: 2146-3042, DOI: 10.25095/mufad.536083, (2019)
17. Filipa Santos, Rúben Pereira and José Braga Vasconcelos: “Toward robotic process automation implementation: an end-to-end perspective”, Business Process Management Journal, ISSN: 1463-7154, (2019/09/30)
18. Chris Lamberton, Damiano Brigo, Dave Hoy: “Impact of Robotics, RPA and AI on the Insurance Industry: Challenges and Opportunities”, Journal of Financial Perspectives, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3079495, (2017)
19. Hussein A. Abbass: “Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-Autonomy Trust”, Springer, https://link.springer.com/article/10.1007/s12559-018-9619-0, (2019/01/14)
20. Nagesh Deshmukh: “Revolutionizing Business Efficiency: The Synergy of RPA and AI, 2023”, https://www.linkedin.com/pulse/revolutionizing-business-efficiency-synergy-rpa-ai-nagesh-deshmukh, (2023/05/21)

Exploring Possibilities of Effectiveness for Integration of AI in High Schools Teaching: Teachers Point of View

Rezak Jakupi[1] and Neroida Selimi[2]

[1] Dardania College, Prishtina 10000, Kosovo
[2] South East European University, Ilindenska 335, Tetovo 1200, North Macedonia
neroida.selimi@seeu.edu.mk

Abstract. This research analyses the potential effectiveness of integrating artificial intelligence (AI) in high school teaching from the perspective of teachers. With AI emerging as a transformative force in education, it is essential to understand teachers' perceptions, attitudes, and experiences regarding its integration into teaching practices. The research aims to explore the opportunities, challenges associated with AI integration in high school education, focusing specifically on the viewpoints of teachers. By using a qualitative approach, through focus groups conducted with high school teachers from Pollog and Skopje region, North Macedonia, the research collects data regarding their perception of AI integration. The qualitative data are analyzed thematically to identify common themes, with the purpose of simplifying and unifying the outcomes. The findings reveal a nuanced understanding of teachers' perspectives on AI integration in high school teaching. While many teachers recognize the potential benefits of AI, such as personalized learning, efficiency enhancement, and opportunities for innovative teaching methods, they also express concerns about challenges such as, data privacy, and the need for comprehensive training and support. This research identifies key factors influencing teachers' attitudes towards AI integration, including familiarity with AI technologies, pedagogical beliefs, confidence in technological skills, and concerns about ethical implications. These insights contribute to a deeper understanding of the complexities surrounding AI integration in high school teaching and offers practical recommendations for enhancing its effectiveness while addressing teachers' concerns and ensuring ethical and responsible use.

Keywords: teaching, artificial intelligence, high schools, teachers, effectiveness.

1 AI, Teaching and Teachers

Artificial intelligence (AI) is used and integrated in every sector. The primary lead sectors on the integration of AI are in healthcare and business [1]. The following sectors are information technology (IT) and education, next being automotive, transportation, telecommunications, manufacturing, etc. [1]. In education, AI makes teaching and

learning more effective by using a set of tools, resources and applications by both parties involved: teachers and students [2].

But is AI used only because of these three technological components; tolls, resources and applications; or is there more? The actual benefits of implementing AI in teaching process are various. Personalized learning is the main feature that AI enables in teaching, followed by increased efficiency, improved feedback, better accessibility, improved data analysis, global collaboration, and emerging technologies [3]. Personalized learning, or tutoring one – on – one has been proven to impact the students' performance for better compared to ones tutored in traditional educational methods [4]. Because of the high costs and large number of teachers needed to implement personalized learning, this has been postponed. With integration of AI in education, personalized learning has been implemented by using adaptive learning platforms to dynamically adjust content and pace based on student performance, creating individualized learning plans with customized materials and activities, and continuously assessing student progress, with the aim to create more student-centered, adaptive, and effective learning experiences, further promoting deeper understanding, engagement, and achievement for every student individually [5]. Increased efficiency is achieved through automated administrative and repetitive tasks, such as grading, creating schedules, managing student data, etc. [6]. Improved feedback is achieved through personalized and instant feedback to students' work, which additionally contributes to identification of mistakes, clarifying concepts and corrections in real – time, leading to faster learning and skill acquisition [7]. When it comes to accessibility, AI helps to address accessibility issues by providing alternative formats for learning materials, such as text-to-speech or speech-to-text capabilities, making education more inclusive for students with diverse learning needs, location variations; such as online or in - class educations, etc. [8]. In the case of data analysis, AI can analyze large volumes of student data to identify patterns, trends, and areas where students may be struggling enabling teachers to intervene early and provide additional support to students [9]. The global collaboration through AI in education enables collaboration and communication among students/ teachers/ schools from different geographical locations, enabling them to work together on projects and exchange ideas in real-time [10]. Emerging technologies, as the phrase says itself, allows students to explore and experiment with cutting-edge technologies, enforcing development, innovation and creativity [11].

Even though there are plenty of benefits that AI brings to education, there are also some disadvantages to it. Main ones are privacy and data security, algorithm bias, overreliance on technology, loss of human connection, cost and resource issues, lack of pedagogical skills [12]. Privacy and data security are the first and main concern because AI systems in education often collect and analyze large amounts of student data, making it possible for sensitive student information to be compromised or misused, especially if proper safeguards and encryption measures are not in place [13]. Algorithmic bias is created based on the data that has been used to train AI algorithms, which can further amplify the existing biases present in the educational materials [14]. Example of algorithmic bias can be seen if the AI models are reflecting certain gender stereotypes, they may produce biased outcomes affecting certain groups of students. In the case of losing

human connections, the impact is through the reduced amount of direct interaction between students and teachers, which diminishes the importance of human relationships and mentorship in education [15]. The cost and resource issues are mostly of a financial nature, because purchasing new technology is costly; and also training the teachers to use and implement the new technology has a significant cost, but otherwise it cannot be used properly [16]. Lack of pedagogical skills refers to the educational staff or teachers and their ability on how to effectively implement AI into their teaching process, or how to interpret the insights provided by the AI systems [17].

With all the benefits and disadvantages that AI can bring to teaching, integrating AI in teaching is an inevitable process. This mainly because of two reasons: AI is integrated everywhere [18]; and the effects of the benefits of AI in education outweigh the effects of the disadvantages [19].

1.1 Literature review

Though AI is a topic that firstly emerged during 1956, by computer scientist John McCarthy [20], its integration in education began during 1970. The earliest examples of AI in education are: intelligent tutoring systems (ITS); computer – based learning programs; educational software and tools; and learning analytics and data mining [21].

What has made AI to thrive in education? There are several AI characteristics that make its integration in education crucial. First of them is personalization. AI can personalize learning experiences for students by adapting content, pace, and teaching methods to individual needs and learning styles [22]. Next one is adaptability. AI adapts to changes in student performance, preferences, and requirements over time, because it continuously learns and improves based on feedback and data [23]. Efficiency is a characteristic of AI in education obtained through automation of routine tasks such as grading assignments, generating quizzes, and organizing course materials, which results in saving teachers time and allowing them to focus on other activities such as lesson planning and student support [6]. Accessibility is enabled by providing alternative formats for learning materials, such as audio descriptions or text-to-speech functionality, to meet different learning needs and preference [8]. AI can analyze large amounts of educational data, including student performance, engagement metrics, and learning patterns, to identify trends, patterns, and insights that can inform instructional decision-making and improve educational outcomes [24]. Adaptive learning systems that have AI integrated, can adjust the difficulty level and content of learning activities based on individual student performance, ensuring that each student is appropriately challenged and supported [25]. AI provides support for teachers by giving them real-time analytics, data insights, and recommendations to better their instructional practices, identify at-risk students, and personalize interventions [26]. AI technologies such as natural language processing (NLP) can facilitate communication and interaction between students and educational systems through voice recognition, chatbots, virtual tutors, and intelligent tutoring systems [27]. AI learning platforms provide 24/7 access to educational resources and support services anytime, anywhere, allowing students to engage in

learning activities at their own pace and convenience [12]. AI enhances student engagement and motivation by incorporating interactive elements, gamification, and personalized learning pathways that cater to students' interests and preferences [28].

The countries that are most developed and have tech giants as home companies, are also the ones that thrive in using AI in every sphere, not just education. Top five worldwide countries that have implemented AI in education are United States of America (USA), followed by China, United Kingdom (UK), Israel, and Canada [29]. France, India, Japan, Germany and Singapore are the five other countries on this list. Educational institutions across Europe are progressively integrating AI initiatives for personalized learning. However, this integration is undertaken cautiously, with careful consideration of its impact on student data privacy and adherence to ethical principles in AI utilization [16].

Most of the countries mentioned above are considered to be first world countries. This means that these are the most developed and industrialized countries with capitalist economies [30]. These countries can afford to invest in AI, and specifically to implement it in their educational systems. What happens with the third world countries, which are considered as developing countries with lower economic performance, such as Western Balkan countries (North Macedonia, Serbia, Kosovo, Albania, Montenegro, Bosnia and Herzegovina)? As the cost for AI in general is high, such is also the cost of maintaining it, training the educators, and also dealing with the disadvantages it brings [31]. Successful models of digital transformation in education are implemented by domestic IT companies in the Western Balkans, which utilize specially designed platforms at the national level, facilitating online teaching, student monitoring, communication, and organizational management through ERP solutions [31]. In this way the cost for AI in education is decreased and more manageable.

Needs for reform and performance analysis of the existing educational systems before the AI integration in education is addressed through Ilic et all. research done [32]. Their research focuses on identifying necessary improvements for the integration of AI technologies in higher education strategy, considering their potential benefits. This research indicates that artificial intelligence and machine learning can enhance skill development, collaborative learning, and research accessibility in higher education institutions.

Kraja focuses on smart education reports [33], which provide insight to the digital transformation of education in both K-12 and higher education levels. The report evaluates the level of digitization in the education system, discusses policies aimed at fostering this process, and outlines key practices implemented at various levels to promote intelligent education. Despite some progress, such as investments in broadband networks and the establishment of agencies and educational centers for AI learning, challenges remain. The analysis highlights the need for increased government investment in education to support smart education initiatives. The report identifies areas for improvement, including digitization efforts, integration of innovative technologies, development of digital resources, and teacher training.

With a subject being developed and researched about for more than 50 years, such as the case of AI, one could easily assume there is a large number of data and literature available. Yet, in the case of AI in education, and specifically in examining teachers'

points of view for integration in teaching in North Macedonia, there is a data and literature gap. For this reason, doing this research, will contribute to the existing gap and enrichment of the data and literature.

Following parts, starting from research proposition, state the questions that this research answers. The methodology part consists of the methodology used, targeted sample, and data collection. Finding describe the analysis emerged from the collected data. The conclusion gives an answer to the research propositions. The last part of this research paper, limitations and future works, states the obstacles and possibilities for further research.

1.2 Research Propositions

This research paper aims to answer the following questions based on the above information:

RQ1. "Do teachers in high schools perceive AI integration as potentially effective in enhancing student learning outcomes, increasing teaching efficiency, and providing opportunities for personalized instruction?"

RQ2. "Do teachers' attitudes and perception vary based on their familiarity with AI, pedagogical beliefs, confidence in their own technological skills, concerns about the ethical implications, and potential drawbacks of AI integration?"

RQ3. "Do teachers need comprehensive training, ongoing support, and clear guidelines for the ethical and responsible use of AI in the classroom to maximize its effectiveness and mitigate potential challenges?"

2 Methodology

The methodology used for this research is qualitative. Qualitative research is a method used to explore and understand people's experiences, perceptions, and behaviors in-depth [34]. Instead of relying on numerical data or statistical analysis, qualitative research gathers descriptive information through techniques such as interviews, focus groups, observations, and textual analysis.

The technique selected for this research is focus group. Focus groups are a qualitative research method used to gather insights and opinions from a small group of participants on a specific topic or issue [35].

2.1 Targeted Sample

Purposive sampling is a non-probabilistic sampling technique used in research to select participants based on specific criteria relevant to the research objectives [36]. Unlike random sampling, which involves selecting participants at random from a population, purposive sampling involves deliberately selecting participants who possess certain characteristics, experiences, or expertise that are of interest to the research study [37].

For this research, a purposive sampling approach is utilized to select participants who have relevant experience and expertise in high school teaching. Participants are

grouped based on years of teaching experience, familiarity with technology, and willingness to participate in the study.

2.2 Data Collection

High schools from North Macedonia are contacted via email during February 2024 to select a set of eligible participants in this research. From the contacted high schools, the ones that were willing to participate in this research are from the Pollog and Skopje region, North Macedonia: Skopje, Tetovo, Negotino (Pollog region), Gostivar. Because of four cities, four focus groups were created by grouping the participants based on the place, experience and technology knowledge.

Focus groups are conducted with selected participants to facilitate open-ended dialogue and group interaction. The focus group discussions are guided by a semi-structured interview designed to explore teachers' perceptions, experiences, and expectations regarding AI integration in high school teaching, guided towards the stated above research propositions. The focus groups are moderated by a trained researcher who facilitated the discussion, encouraged participation, and ensured that all participants have an opportunity to share their perspectives. Each focus group session lasted approximately 60-90 minutes, is audio-recorded with the permission of participants, translated in English language for unification purposes as the focus groups were in Albanian and Macedonian language, further transcribed and analyzed.

All the participants requested to be anonymously stated in the research, for which purpose they will further be referred to the first letter noting the place (S for Skopje, T for Tetovo, N for Negotino, and G for Gostivar), followed by a number; example G1 is participant 1 from Gostivar, S3, is participant 3 from Skopje. This abbreviation is used for simplifying the collected data for further analysis.

For the focus groups, participants are group as stated below:

- Group 1 (G1) – S1, S5, T1, G1, G6, N1, N5
- Group 2 (G2)– S2, S6, T2, T5, G2, N2
- Group 3 (G3)– S3, S7, T3, G3, G5, N3
- Group 4 (G4)– S4, S8, T4, T6, G4, N4

The focus group interviews are conducted through March/ April 2024. In the following sections focus groups are abbreviated with notation G noting group and a number (G1, G2. G3, G4).

3 Findings

This part of the research is separated into four parts: general data, which presents information about the participants such as place, high schools, age of experience and technology knowledge; analysis and data presentation for RQ1; analysis and data presentation for RQ2; and analysis and data presentation for RQ3.

3.1 General Data

Table 1 provides a structured overview of the demographic and professional attributes of the participants involved in the study, allowing for analysis and comparison across different groups. There are a cumulative 8 participants from Skopje region, 6 are from Tetovo region, 6 are from Gostivar region, and 5 are from Negotino region. A total of 25 participants were included in this research. The age varies from 25 to 62 years old, genders male 13 and female 12, experience from 1 to 35 years of experience in education, and technology knowledge from 4 to 10, from a Likert scale 1 being no technology knowledge and 10 being proficient technology knowledge. For this focus group, the technology knowledge refers to the ability of the participants to use computers, internet services, educational software, Microsoft package, etc..

Table 1. General Data

Partic.	High School	Age	Gender	Exp.	Tech. Know.	Subject
S1	HS “Zef Lush Marku”	31	F	5	10	Informatics
S2	HS “Zef Lush Marku”	25	F	1	10	Marketing
S3	HS “Zef Lush Marku”	42	F	15	8	Business
S4	HS “Zef Lush Marku”	48	M	20	8	Business
S5	ELHS “Arseni Jovkov”	52	M	25	5	Sociology
S6	ELHS “Arseni Jovkov”	47	M	20	7	Albanian Lan.
S7	ELHS “Arseni Jovkov”	44	F	20	5	Macedonian Lan.
S8	ELHS “Arseni Jovkov”	50	M	25	10	Informatics
T1	HS “Kiril Pejqinovic”	30	F	3	10	Mathematics
T2	HS “Kiril Pejqinovic”	32	M	1	10	Legal
T3	HS “Kiril Pejqinovic”	44	M	18	7	Albanian Lan.
T4	HS “Kiril Pejqinovic”	41	F	5	9	Sociology
T5	MSH “Nikola Shtejn”	46	F	12	6	Albanian Lan.
T6	MSH “Nikola Shtejn”	53	F	25	4	Macedonian Lan.
G1	HS “Gostivar”	62	M	35	6	Sociology
G2	HS “Gostivar”	45	M	10	6	Business
G3	MHS “Gostivar”	27	F	2	7	Physiotherapy
G4	HTS “Gostivar”	25	M	1	9	Sports
G5	MTS “Gostivar”	50	F	13	5	Nursing
G6	MTS “Gostivar”	29	F	4	4	Physiotherapy
N1	HS “Naser Ademi”	38	M	13	10	Business
N2	HS “Naser Ademi”	32	M	2	7	Sports
N3	HS “Naser Ademi”	57	F	30	4	Physics
N4	HS “Naser Ademi”	42	F	7	7	Music
N5	HS “Naser Ademi”	55	M	30	6	Arts

From the above table (Table 1) can be concluded that the majority of participants with proficient technology knowledge are younger or with IT background. The majority of the participants with less technology knowledge are older or with background different from IT.

3.2 Analysis RQ1

For the first proposition, there are discussed the following points when it comes to teachers for AI integration in education: enhanced student learning outcomes, increased teaching efficiency, and providing opportunities for personalized instruction. Noted in the brackets are the groups supporting the statements.

Enhanced Student Learning Outcomes. The majority of teachers in high schools see AI integration as highly effective in enhancing student learning outcomes (G1, G2, G3, G4). They think that usage of AI tools and resources can provide personalized learning experiences tailored to individual student needs, leading to improved academic performance and engagement. The concern of the teachers in this case is their belief that the cost for AI tools and resources is still high for integration in North Macedonia (G1, G2, G3), and maintenance and training for usage is even costlier (G2, G4).

Increased Teaching Efficiency. Teachers also recognize the potential of AI integration to increase teaching efficiency by automating routine tasks, such as grading assignments and providing feedback (G1, G2, G4). Focus groups mentioned the educational system used during the COVID-19 pandemic, Google Classroom, were they could automatically calculate the grade based on the performance on the assignments. The discussions also noted that automatic grading could not include the part of how active a student was in class, which made them lose points (G2, G3). Also, a problem noted from the participants was that majority of their colleagues, especially the ones with more experience in teaching, had problems in using the automatic grading, need more instructions and training, and favored more their own grading policies (G1, G2, G3, G4).

Opportunities for Personalized Instruction. The participants in this study did not have the opportunity to use systems that provide data analysis personalized to student's needs, but they agree that having this kind of insight would enable differentiated instruction, allowing them to tailor learning experiences to each student's strengths, weaknesses, and interests (G1, G2, G3, G4).

3.3 Analysis RQ2

The second proposition analysis are teachers biased towards AI in education based on their own abilities to use it and technological knowledge through following points: familiarity with AI, pedagogical beliefs, confidence in their own technological skills, and their concerns about AI drawbacks. Noted in the brackets are the groups supporting the statements.

Familiarity with AI. Teachers' attitudes and perceptions towards AI integration vary based on their familiarity with AI technologies (G1, G2, G3, G4). Teachers who

are more familiar with AI tend to have more positive attitudes towards its integration in education. Contrary to this, teachers who are less familiar with AI are more sceptic about its effectiveness and practicality in the classroom.

Pedagogical Beliefs. Teachers' pedagogical beliefs influence their attitudes towards AI integration (G1, G2, G3, G4). Teachers who hold constructivist or student-centered pedagogical beliefs may view AI as a valuable tool for facilitating student-centered learning experiences and promoting active engagement (G1, G4). On the other hand, teachers with more traditional or teacher-centered pedagogical beliefs may be more cautious to AI integration, preferring more traditional instructional methods and approaches (G2, G3, G4).

Confidence in their own Technological Skills. Teachers' confidence in their own technological skills affects their attitudes towards AI integration (G1, G2, G3, G4). Teachers who are confident in their technological abilities or have more technological knowledge, may embrace AI integration as an opportunity to enhance their teaching practices and benefit from technology to support student learning (G1, G2, G3). The participants noted that teachers who lack confidence in their technological skills, have a background that does not relate at all to having IT skills, or are older, may feel overwhelmed by AI integration, expressing concerns about their ability to effectively use and integrate AI technologies into their teaching (G3, G4).

Concerns about AI Drawbacks. Teachers express concerns about the ethical implications and potential drawbacks of AI integration in education (G1, G2, G3, G4). Even though the participants do not use any AI systems that provide data analysis, their main concern is towards data privacy and security related to the collection and use of student data by AI systems (G1, G3, G4). Additionally, teachers are concerned about AI replacing human judgment, empathy, and connection, as the case they see with usage of mobile phones, which additionally impacts the attention span and psychological health of the students (G1, G2, G3).

3.4 Analysis RQ3

The third proposition analyses teachers need to use AI in the best way possible through the following points: comprehensive training, ongoing support, clear guidelines. Noted in the brackets are the groups supporting the statements.

Comprehensive Training. Teachers express a strong need for comprehensive training in AI technologies to effectively integrate them into their classroom practices (G1, G2, G3, G4). Many teachers acknowledge that they lack the necessary knowledge and skills to use AI tools and resources confidently, and in their opinion these training should be done before the integration of AI in education (G1, G2, G3, G4). Participants noted that these training should be tailored to their specific subject areas, grade levels, and teaching contexts to ensure relevance and applicability (G1, G4). There was also commented that teachers that are near pensions, probably do not need these trainings, as in their opinions they would not benefits that much (G1, G3).

Demand for Ongoing Support. The participants agree noted that teachers need an ongoing support and professional development opportunities to sustain their use of AI in the classroom over time (G1, G2, G3, G4). They stated that learning about AI is an

ongoing process that requires continuous support and access to resources, same as their preparations for lectures and teaching (G2, G4). Additionally, participants noted that teachers need the technical support and troubleshooting assistance to address issues and concerns that may arise during the implementation of AI technologies in their classrooms, even after the trainings (G1, G3, G4). They also stated that this should be done by IT support professionals, who would be available for assistance during their scheduled lectures.

Clear Guidelines. Teachers express a need for clear guidelines and ethical frameworks for responsible use of AI in the classroom (G1, G2, G3, G4). The participants stated concerns about ethical considerations such as data privacy, security, bias, and transparency in AI systems (G1, G3). Teachers emphasize the importance of clear guidelines that outline best practices, ethical standards, and legal requirements for the collection, storage, and use of student data (G1, G2, G3, G4). They noted the need for guidelines that address issues such as informed consent, data protection, algorithmic transparency, and accountability to ensure the ethical and responsible use of AI in education (G1, G2, G3, G4).

4 Conclusion

In this part of the research paper can be found answers to our research questions mentioned in the research propositions.

RQ1: "Do teachers in high schools perceive AI integration as potentially effective in enhancing student learning outcomes, increasing teaching efficiency, and providing opportunities for personalized instruction?"

The majority of high school teachers perceive AI integration as highly effective for better student learning outcomes. The participants in this research believe that AI tools and resources can provide personalized learning experiences tailored to individual student needs, leading to improved academic performance and engagement. However, teachers' express concerns about the cost of AI tools and resources, as well as the expenses associated with maintenance and training for usage. Additionally, the participants stated that teachers recognize the potential of AI integration for increased teaching efficiency by automating routine tasks such as grading assignments and providing feedback. They also note challenges with automatic grading systems, including difficulty in capturing students' class participation and resistance from colleagues who require more training and favor their own grading policies. While the participants did not have the opportunity to use systems that provide personalized data analysis, they agree that having such insights would enable differentiated instruction tailored to each student's strengths, weaknesses, and interests.

RQ2: "Do teachers' attitudes and perception vary based on their familiarity with AI, pedagogical beliefs, confidence in their own technological skills, concerns about the ethical implications, and potential drawbacks of AI integration?"

Teachers' attitudes towards AI integration vary based on their familiarity with AI technologies. Those who are more familiar tend to have more positive attitudes, while those who are less familiar are more skeptical about its effectiveness and practicality in

the classroom. Similarly, teachers' pedagogical beliefs influence their views on AI integration. Those who hold constructivist or student-centered beliefs see AI as a valuable tool for promoting active engagement, while those with more traditional beliefs may be more cautious. Additionally, teachers' confidence in their own technological skills plays a role in their attitudes towards AI integration. Those who are confident or have more technological knowledge may embrace AI as an opportunity to enhance teaching practices, while others may feel overwhelmed and express concerns about their ability to effectively use AI technologies. Furthermore, participants stated that teachers express concerns about the ethical implications and potential drawbacks of AI integration, particularly regarding data privacy and security. They also worry about AI replacing human judgment and connection, impacting students' attention span and psychological health.

RQ3: "Do teachers need comprehensive training, ongoing support, and clear guidelines for the ethical and responsible use of AI in the classroom to maximize its effectiveness and mitigate potential challenges?"

Teachers express a strong need for comprehensive training in AI technologies to effectively integrate them into their classroom practices. Many acknowledge their lack of necessary knowledge and skills, emphasizing the importance of training before AI integration. They suggest that training should be tailored to specific subject areas, grade levels, and teaching contexts to ensure relevance and applicability. However, some suggest that teachers near retirement may not benefit as much from these trainings. Additionally, participants agree that ongoing support and professional development opportunities are essential to sustain AI use in the classroom over time. They emphasize the need for continuous support and access to resources, as well as technical support and troubleshooting assistance during implementation. Participants suggest that IT support professionals should be available for assistance during scheduled lectures. Moreover, teachers express a need for clear guidelines and ethical frameworks for the responsible use of AI in the classroom. They raise concerns about ethical considerations such as data privacy, security, bias, and transparency in AI systems. Teachers stress the importance of clear guidelines outlining best practices, ethical standards, and legal requirements for the collection, storage, and use of student data. They highlight the need for guidelines addressing issues such as informed consent, data protection, algorithmic transparency, and accountability to ensure the ethical and responsible use of AI in education.

4.1 Emerging Propositions

From the conclusions stated in the above section, the following propositions emerge:

- Promoting and supporting AI integration: Promoting the use of AI for personalized learning and improving student outcomes by providing financial support for AI tools and teacher training.
- Customized teacher training: Developing tailored training programs to increase teachers' familiarity and confidence with AI, ensuring effective classroom integration.

- Ongoing support: Establish continuous professional development and IT technical support to sustain AI use in schools, by facilitating the communication between the teachers and IT support professionals.
- Ensuring ethical AI use: Developing clear guidelines for ethical AI implementation, addressing concerns like data privacy, algorithmic bias, and transparency.
- Adapting AI to diverse teaching approaches: Ensuring AI tools are flexible enough to support both traditional and student-centered teaching methods, considering teachers various pedagogical beliefs.

5 Limitations and Future Works

Limitations of this research are geographical. The research focuses on high school teachers from Pollog and Skopje region, North Macedonia. Even though schools from different municipalities of North Macedonia have been contacted, most of them haven't responded to the request, and some have not accepted to participate in the study.

In the case of the study expansion with a larger number of participants, consider the qualitative methodology included, which shifts the used methodology to a mixed one, as more appropriate for in depth insight.

The teaching process consists of several main actors, such as the government and ministry for education and science, the management or board of the educational facility, the teachers, administrators, students, parents. This research only tackles the point of view of teachers. The perspective of other actors can also be researched. This would enable a bigger data set, testing of more propositions and hypothesis, and having a multiple point of view about this topic.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. H. Ming-Hui and R. Rust, "Artificial Intelligence in Service," *Journal of Service Research,* vol. 21, no. 2, pp. 155-172, 05 February 2018.
2. R. Luckin, K. George and M. Cukurova, AI for School Teachers, vol. 1st, Boca Raton: CRC Press, 2022, p. 132.
3. Z. Budee, "Transforming Education Through AI Benefits Risks and Ethical Considerations," *TechRxiv,* 05 October 2023.
4. E. D. Brown, Exploration of a Pedagogical Shift: Transitioning from Traditional Teaching to Personalized Learning, 2020.
5. T. van der Vorst and N. Jelicic, "Artificial Intelligence in Education: Can AI bring the full potential of personalized learning to education?," *30th European Conference of the International Telecommunications Society (ITS): "Towards a,* 19 June 2019.

6. L. Chen, P. Chen and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access,* vol. 8, pp. 75264-75278, 2020.
7. R. Luckin and W. Holmes, Intelligence Unleashed: An argument for AI in Education, London: Pearson, 2016.
8. A. Goel, "AI-Powered Learning: Making Education Accessible, Affordable, and Achievable," 2020.
9. S. J. H. Yang, "Guest Editorial: Precision Education - A New Challenge for AI in Education," *International Forum of Educational Technology & Society,* vol. 24, no. 1, pp. 105-108, January 2021.
10. M. Bottery, "Education and globalization: redefining the role of the educational professional," *Educational Review - Routledge,* vol. 58, no. 1, pp. 95-113, 16 August 2006.
11. T. Wang and E. C. K. Cheng, "Towards a Tripartite Research Agenda: A Scoping Review of Artificial Intelligence in Education Research," *Artificial Intelligence in Education: Emerging Technologies, Models and Applications,* 2022.
12. B. A. Chaushi, F. Ismaili and A. Chaushi, "Pros and Cons of Artificial Intelligence in Education," *International Journal of Advanced Natural Sciences and Engineering Researches,* vol. 8, pp. 51-57, 01 March 2024.
13. A. Kamenskih, "The analysis of security and privacy risks in smart education environments," *Journal of Smart Cities and Society,* vol. 1, no. 1, pp. 17-29, 14 February 2022.
14. R. Baker and A. Hawn, "Algorithmic Bias in Education," *International Journal of Artificial Intelligence in Education,* vol. 32, pp. 1052-1092, 18 November 2021.
15. K. Seo, J. Tang, I. Roll, S. Fels and D. Yoon, "The impact of artificial intelligence on learner–instructor interaction in online learning," *International Journal of Educational Technology in Higher Education,* vol. 18, no. 54, 26 October 2021.
16. F. Tahiru, "AI in Education: A Systematic Literature Review," *Journal of Cases on Information Technology (JCIT),* vol. 23, no. 1, 2021.
17. S. Pokrivcakova, "Preparing teachers for the application of AI-powered technologies in foreign language education," *Journal of Language and Cultural Education,* vol. 7, no. 3, pp. 135-153, 26 December 2019.
18. G. Lew and R. Schumacher, "AI-Enabled Products Are Emerging All Around Us: Technology is everywhere," pp. 55-85, October 2020.
19. A. Guilherme, "AI and education: the importance of teacher and student relations," *AI & Society,* vol. 34, pp. 47-54, 2019.
20. "History of artificial intelligence," Wikipedia, 2024. [Online]. Available: https://en.wikipedia.org/wiki/History_of_artificial_intelligence.
21. B. Williamson and R. Eynon, "Historical threads, missing links, and future directions in AI in education," *Learning, Media and Technology,* vol. 45, no. 3, pp. 223-235, 2020.
22. W. Holmes and I. Tuomi, "State of the art and practice in AI in education," *European Journal of Education,* vol. 57, no. 4, pp. 542-570, 30 October 2022.
23. S. Wang, C. Christensen, W. Cui, R. Tong, L. Yarnall, L. Shear and M. Feng, "When adaptive learning is effective learning: comparison of an adaptive learning

system to teacher-led instruction," *Interactive Learning Environments,* vol. 31, pp. 793-803, 31 August 2020.

24. N. Humble and P. Mozelius, "Teacher-supported AI or AI-supported teachers?," *European Conference on the Impact of Artificial Intelligence and Robotics,* pp. 157-164, 2019.
25. K. Fuchs, "Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse?," *Frontiers in Education,* vol. 8, 2023.
26. W. Neji, N. Boughattas and F. Ziadi, "Exploring New AI-Based Technologies to Enhance Students' Motivation," *Issues in Informing Science and Information Technology,* vol. 20, pp. 95-110, 2023.
27. T. Keary and E. Wrenn, "Top 10 Countries Leading in AI Research & Technology in 2024," Technopedia, 09 April 2024. [Online]. Available: https://www.techopedia.com/top-10-countries-leading-in-ai-research-technology.
28. C. Banton, P. Westfall and T. Li, "Third World Countries: Definition, Criteria, and List of Countries," Investopedia, 23 February 2024. [Online]. Available: https://www.investopedia.com/terms/t/third-world.asp.
29. M. Travar, I. Dugonjic, Z. Avramovic, G. Bajic and S. Ristic, "Digital Transformation of the Education Sector in the Western Balkans," *The 1st International Conference on Maritime Education and Development,* pp. 183-190, 25 March 2021.
30. M. P. Ilic, D. Paun, N. P. Sevic, A. Hadzic and A. Jianu, "Needs and Performance Analysis for Changes in Higher Education and Implementation of Artificial Intelligence, Machine Learning, and Extended Reality," *Education Sciences,* vol. 11, no. 10, 2021.
31. P. Kraja, "Report on Smart Education in Albania," *Smart Education in China and Central & Eastern European Countries,* pp. 51-79, 2023.
32. A. Sinha and K. Majumder, RESEARCH METHODOLOGY (A Guide for Scholars), Pune: Kripa-Drishti Publications, 2021.
33. V. A. Wilson, "Focus Groups: a useful qualitative method for educational research?," *British Educational Research Journal,* vol. 23, pp. 209-224, 1997.
34. J. M. Guarte and E. B. Barrios, "Estimation Under Purposive Sampling," *Communications in Statistics - Simulation and Computation,* vol. 35, pp. 277-284, 2006.
35. A. S. Acharya, A. Prakash, P. Saxena and A. Nigam, "Sampling: why and how of it?," *Indian Journal of Medical Specialities,* vol. 4, no. 2, pp. 330-333, 2013.
36. D. C. Moos and R. Azevedo, "Learning With Computer-Based Learning Environments: A Literature Review of Computer Self-Efficacy," *Review of Educational Research,* vol. 79, no. 2, pp. 576-600, 2009.
37. D. Kaminska, G. Zwolinski, H. Maloku, M. Ibrani, J. Guna, M. Pogacnik, R. E. Haamer, G. Anbarjafari, L. A. Bexheti, K. Bozhiqi and A. Halili, "The Trends and Challenges of Virtual Technology Usage in Western Balkan Educational Institutions," *Information,* vol. 13, no. 11, 2022.

Scalability and Performance of a Custom-made RESTful API

Igor Janevski[1*], Marjan Gushev[1*] and Nevena Ackovska[1*]

[1*]Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Skopje, North Macedonia.

*Corresponding author(s). E-mail(s): igor_janevski@hotmail.com; marjan.gushev@finki.ukim.mk; nevena.ackovska@finki.ukim.mk;

Abstract

Web services are used more than ever before due to their ability to simplify communication with third-party applications or other web services through APIs. However, many web services do not perform well under increasing load, requiring a thorough analysis of their scalability and performance. This research paper evaluates the scalability and performance of a custom-made RESTful web service, specifically a Weather API. We perform performance testing on this web service to identify and discuss bottlenecks and other performance issues and provide potential solutions.

Keywords: web service deployment, performance testing, scalability, REST API

1 Introduction

This paper aims to test and compare the scalability and performance of a custom-made Weather API web service that uses RESTful (representational state transfer) [1] services for interaction with third-party applications. This Weather API web service consists of a NodeJS [2] application connected to a MongoDB [3] database for data storage, all containerized. It supports standard CRUD (create, read, update, delete) operations of weather records such as air temperature and humidity. The data was collected using a digital digital humidity and temperature (DHT) sensor [4]. To test the performance of this RESTful web service, we plan to send requests (simulating multiple users) constantly and simultaneously and fetch content. To understand potential bottlenecks and other performance issues, experiments include configurations with

different numbers of instances (containers) of the NodeJS application and MongoDB database tests.

The RESTful API (application programming interface) uses RESTful web services to facilitate communication between endpoints, primarily using the JSON file format. A web service represents a software resource (e.g., an application) providing services to different systems or end-users. Communication with web services is performed using an API, which enables interaction with the web service. The most common standards for web service communication are SOAP (simple object access protocol) and the RESTful architectural style. The SOAP protocol uses the XML (extensible markup language) format, while REST primarily uses JSON, although other file formats are also supported.

Scalability is a property of a system that dynamically or manually adjusts the performance of the computing resources according to the needs of the web service [5, 6]. Web services can scale vertically or horizontally. Vertical scalability refers to adding or removing more resources (compute or storage) to existing infrastructure. Horizontal scalability refers to adding or removing resources (such as servers and containers) to share the workload by balancing the shared resources.

Regarding data collection, a web service can use any database, such as MongoDB, used in our use case as a distributed NoSQL [7] database for unstructured and structured data. It is a document-oriented database that uses JSON (or other types of) documents. NodeJS is a server-side technology based on the JavaScript scripting language. The NodeJS application is connected to a MongoDB database. Both are deployed as a web service with an exposed RESTful API; therefore, the term web service will represent the abstracted solution.

The Weather API web service solution is deployed on a local server (on premise) with installed Docker [8], and Kubernetes [9] platforms. We created Docker images for the NodeJS and MongoDB implementations. Kubernetes is an orchestration tool that allows the manipulation of instances of the NodeJS application and the MongoDB database, called pods. These pods can be easily created/destroyed to fit the testing scenarios. Usually, if there are multiple pods of the same type, they are balanced using a load balancer, and we use MetalLB.

The objective of this research paper is to determine whether the web service exhibits linear scalability with increasing load (number of users/requests) and to assess the potential impact of this load on the performance of the web service in various scenarios. Furthermore, the research seeks to identify possible bottlenecks and other issues that could negatively affect the performance of the web service. In summary, this paper explores how well a simple, custom-built Weather API web service can work based on the number of initiated requests (load) and what aspects need to be scaled (improved).

This paper is organized in the following structure. Section 2 presents the related work. The custom-built Weather API is explained in Section 3, along with the experimental and evaluation methods. Section 4 presents and discusses the results. Finally, Section 5 presents the conclusions and future work.

2 Related work

Jun Hong et al. [10] compares the performance of a RESTful API with the message broker RabbitMQ. Interestingly, when the load on the RESTful API was below a certain threshold, it outperformed RabbitMQ. However, as the load surpassed this tipping point, the RESTful API's performance deteriorated, and the error rate increased significantly. In contrast, the RabbitMQ messaging method remained stable under the increased load. This behavior of the RESTful API is similar to what we observed in our testing.

Choi presented another related approach [11], examining the scalability of RESTful API web services and evaluating their performance on mobile and cloud platforms. The author analyzes key performance metrics such as response time, throughput, error rate, and availability, which are crucial for a solid understanding of RESTful API performance. These metrics are considered when performing the performance testing of the Weather API. The author also outlines a specific approach, including benchmarking and comparative analysis of different frameworks and environments, revealing significant variations based on the chosen technologies.

Isha et al. [12] discusses the challenges and solutions of automating API testing. The study highlights how automated testing can significantly reduce the time, cost, and manpower required compared to manual testing. The authors emphasize the importance of addressing issues such as sequencing API calls and managing unpredictable JSON responses. They propose an automated testing tool that improves accuracy and efficiency in detecting errors through automated API calls and response comparisons. The study concludes that automated API testing is essential for enhancing software quality and reducing the regression testing workload. Some of the ideas presented in this article are considered when evaluating the testing methods of the Weather API.

3 Methods

3.1 Use Case Application

All tests are performed on the NodeJS application and MongoDB database. The NodeJS application and MongoDB database are containerized using Docker and Kubernetes. This allows the containers or pods to be easily created or destroyed. MetalLB load balancer balances the pods (Fig. 1). The pods' resources are limited to 20% per logical core on the physical server.

The host machine (from which the tests are run) has an Intel i7 (8 cores) CPU, 16GB RAM, and Windows 10 OS. The server (where the tests are being performed) is a Dell Edge R200 with an Intel Xeon (4 cores) CPU, 8GB RAM, and Ubuntu 21 OS. The network connection between the host and server machine is a 100Mbps LAN (local area network). Moreover, the server is running NodeJS version 16 and MongoDB version 4.4. The tests are performed using JMeter version 5.4.3. These specifications are sufficient for the testing scenarios.

JMeter [13] is a testing tool for measuring performance and analyzing web services and websites used on the Weather API web service. Configuring JMeter correctly is

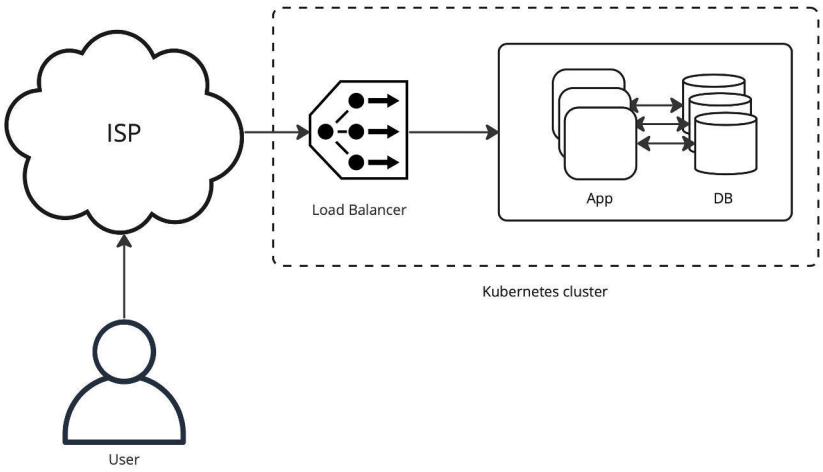


Fig. 1 High-level logical design of the web service.

essential to obtain accurate results. Below are the most critical points and properties for proper load testing in JMeter—it is crucial to have the environment set correctly.

3.2 Test Cases

Three scenarios were created to test the Weather API web service's performance and determine which offers the best results. In the first scenario, three NodeJS applications (pods) are created and connected to a replica set of three MongoDB pods (Fig. 2, Scenario 1). These three NodeJS applications use a round-robin load balancer to distribute the traffic equally. In the second scenario, a single NodeJS application is connected to a cluster with a replica set of three MongoDB pods (Fig. 2, Scenario 2). In the third scenario, three NodeJS applications, each with a round-robin load balancer, are connected to a single MongoDB pod (Figure 2, Scenario 3). An additional scenario is created to determine how many requests are needed to cause the web service to drop some requests or even crash.

The following items define the scenarios and test cases:

- The number of threads - This is the number of concurrent simulated users. From previous dry runs, this number is set to 7000.
- The ramp-up period (in seconds) - This is the time it takes to start all threads (users). In this case, it is set to 100 seconds.
- Specify thread lifetime - This is the operation time of the threads. In this case, it is set to 100 seconds.
- The startup delay time - This is set to 10 seconds.
- Delay thread creation - This is disabled since the host's CPU utilization is always within acceptable limits (under 80%). The CPU utilization should not exceed 80% of the total CPU capacity to prevent bottlenecks in the host machine when performing the tests. If this is not the case, then the host's CPU would not be able to handle

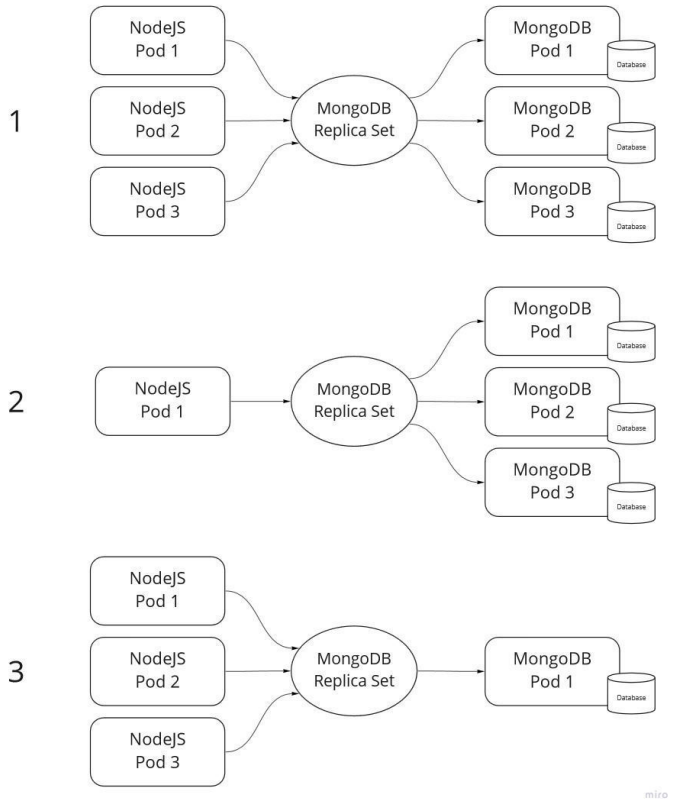


Fig. 2 The image shows the three scenarios that are the focus of this research paper. Scenario 1 involves three containerized NodeJS applications, each connected to one of three database instances configured in a replica set. Scenario 2 features a single containerized NodeJS application connected to three containerized database instances. Scenario 3 includes three containerized NodeJS applications, all sharing a single containerized database instance. The objective is to determine which of these scenarios experiences the fewest bottlenecks and the least performance issues.

the creation of the threads (users) to the full extent, leading to a bottleneck in the host and inaccurate results. In the proposed scenarios, this number is around 20%, much lower than the 80% threshold.

- Cache-Control: no-cache - This parameter is set to ensure that the response will not be cached and reused in future REST requests. Instead, each new REST request to the web service will trigger a new query to the database and send a new response.
- To ensure that there will be no bottlenecks from the Java VM (virtual machine), its heap size is set to accommodate the increased RAM usage when running JMeter. This value is set to 8GB.
- JMeter is executed via the command-line interface (CLI) to ensure the most accurate results.

A mean was calculated for executing each test scenario five times to maintain consistency and minimize fluctuations. The standard deviation for the APDEX value (explained below) was under 0.2, indicating good consistency. No errors were detected during testing, meaning all REST requests were executed successfully. The only exception was the final additional test, which aimed to determine the number of REST requests required to cause the web service's REST API to malfunction and throw errors.

3.3 Evaluation methods

For the tests, it is assumed that the web service always returns the correct data and that there are no issues with its internal logic. When evaluating the performance of the web service, the most commonly considered metric is response time, defined as the time difference between the request and the response. The lower the response time, the better the performance of the web service (assuming that the correct data is returned). Testers typically measure the response time using the Time to First Byte (TTFB) technique [14], which records the time (in milliseconds) it takes for the user to receive the first byte of information from the server.

According to Google's recommendations for RESTful web services, the response time should be under 200ms for a satisfied count (good response), under 1000ms for a tolerating count (acceptable response), and more than 1000ms for an unsatisfied count (delayed response). These timing metrics can vary depending on the web service's purpose and the evaluator's perspective, making it somewhat subjective. Additionally, due to their complexity, some web services may require more time to complete requests, and this longer response time can be considered normal for them.

Based on the information about response time, the application performance index (APDEX) score [15] can be calculated to indicate user satisfaction when using the RESTful web service. The APDEX score is calculated with the following formula:

$$Apdex_T = \frac{Satisfied\ count + Tolerable\ count/2}{Total\ samples} \tag{1}$$

The APDEX index varies between 0 and 1, where values between 1.00 and 0.94 are excellent, between 0.93 and 0.85 are good, between 0.84 and 0.70 are fair, between 69 and 0.49 are poor, and less than 0.49 are bad.

Once the Weather API web service is successfully started and running, Postman [16] tests the initial connection by sending a few REST requests (Fig. 3). Performance testing is the next activity to determine how well the web service can handle multiple users simultaneously. All the simulations are performed with JMeter, simulating users making GET requests to fetch the DHT sensor data. JMeter can generate multiple threads that simulate multiple users making GET requests to the web service API.

4 Results and discussion

After running the tests in JMeter five times for each of the three scenarios and calculating the mean, all REST requests were completed successfully, with no failed requests (except in the particular scenario discussed later). The average result from the five

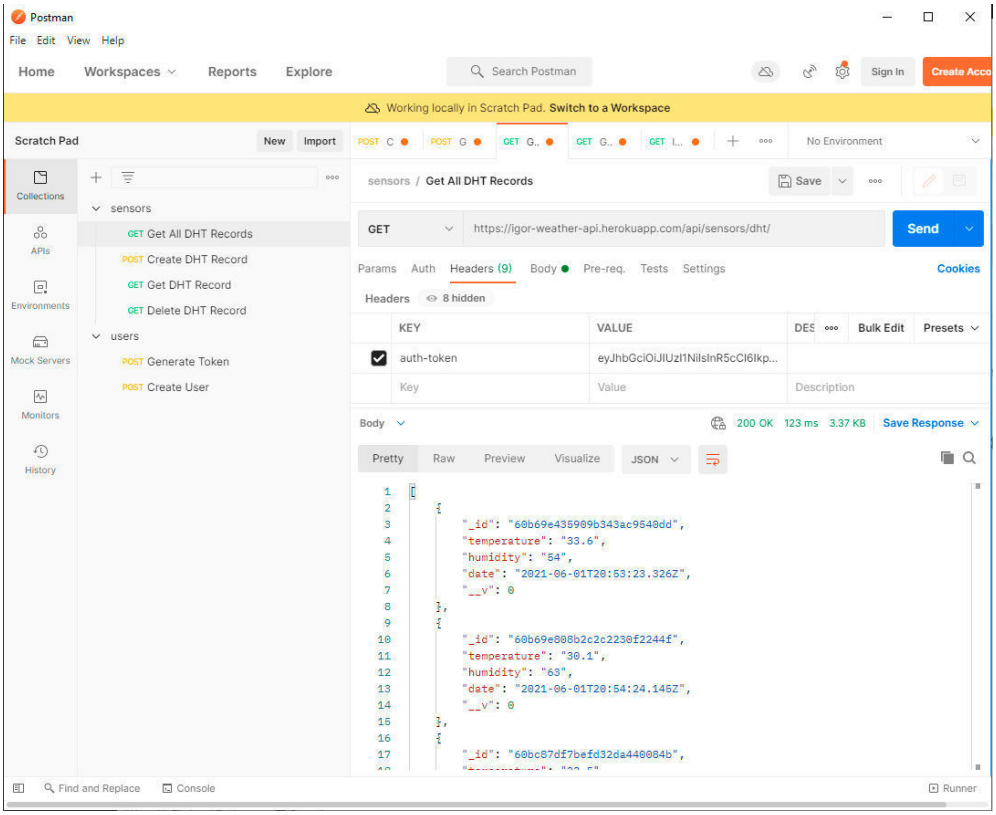


Fig. 3 A sample response from the custom-made Weather API that displays sensor data (temperature and humidity) using Postman.

tests was considered, with a standard deviation of 0.2 for the APDEX index, which falls within the desired range.

The first scenario includes three instances or pods of the NodeJS application with a round-robin load balancer connected to a replica set of three MongoDB instances. The calculated APDEX index is 0.968 (Figure 4), which falls into the excellent category, indicating extremely high user satisfaction. The average response time was 39.90ms, signifying that the requests were processed on time (Figure 5 Graph 1).

The second scenario includes one instance or pod of the NodeJS application connected to a replica set of three MongoDB instances. The corresponding APDEX index is 0.020 (Figure 4), which falls into the poor category, indicating shallow user satisfaction. Although there were no errors in the responses, the average response time was significantly higher at 10,405.75ms (Figure 5 Graph 2).

The third scenario includes three instances or pods of the NodeJS application with a round-robin load balancer connected to a replica set consisting of only one MongoDB instance. The resulting APDEX index is 0.896 (Figure 4), which falls into the good

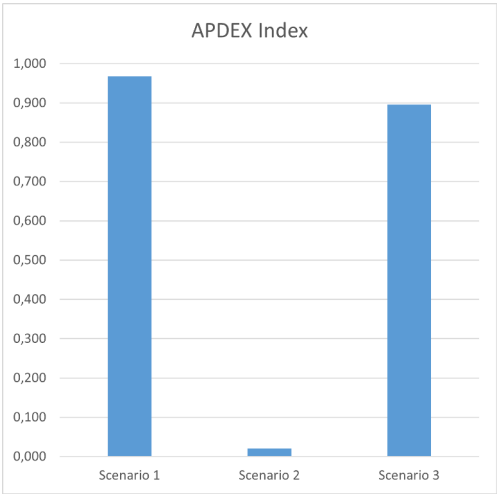


Fig. 4 Calculated APDEX index for each scenario.

category, indicating relatively high user satisfaction. The average response time was 162.84ms (Figure 5 Graph 3).

A particular scenario was created to determine how many REST requests would make the Weather API web service unstable and cause it to drop some requests or crash. The tests revealed that around 8,000 requests were required to induce instability and potentially cause dropped requests, although the web service did not crash.

The results of the three original scenarios show that the APDEX index is high in both the first and third scenarios. Further analysis indicated that the number of NodeJS instances or pods is crucial to the web service's performance, as both scenarios with three NodeJS instances and a round-robin load balancer perform well. Conversely, the second scenario, with only one NodeJS instance, showed a poor APDEX index. Given that other factors are constant in this analysis, it can be concluded that most request processing occurs in the NodeJS application rather than in the MongoDB database. In other words, processing requests in the application is more intensive than processing requests in the database.

When a GET request is sent, the NodeJS application processes it and queries the MongoDB database for results. The result is then returned to the NodeJS application, which is processed again before being returned to the client. The extended response time is likely due to CPU utilization or processing power used by the NodeJS application. In terms of memory, the NodeJS application appeared to have adequate resources.

No issues or bottlenecks were detected in the queries executed on the MongoDB database, confirming that the bottlenecks are occurring in the NodeJS application.

The first and third scenarios achieve stable response times (Fig. 5), whereas the second scenario exhibits a poor response time.

In the particular scenario designed to determine how many REST requests are needed to cause the web service to drop some requests or even crash, it was observed that around 8,000 requests are sufficient to induce request drops. In this case, the

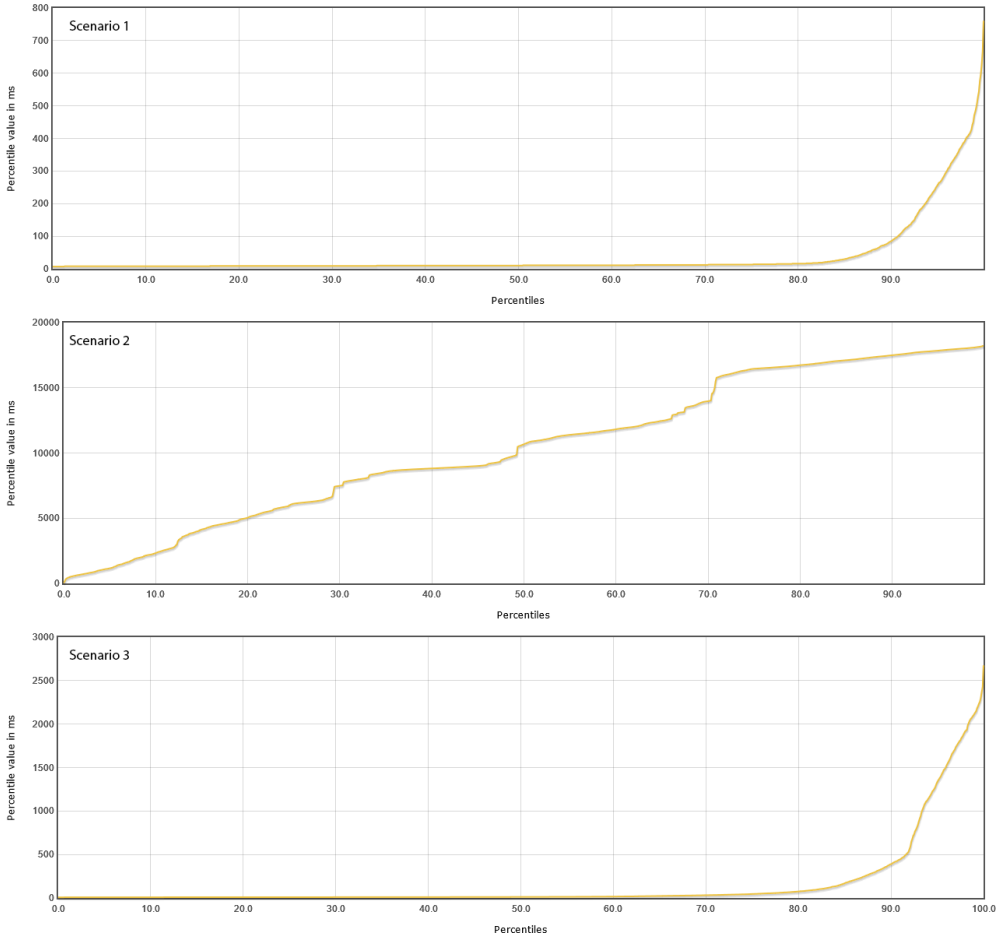


Fig. 5 Performance results for three scenarios, where the X-axis represents the percentile number, while the Y-axis indicates the percentile value (response time) in milliseconds.

APDEX index was not measured; instead, the focus was on the stability of the web service. The dropped requests were attributed to timeouts, indicating that the Weather API web service failed to process some requests. This issue could be due to insufficient memory, such as buffer or heap memory, even though the heap memory was increased to 8GB for testing. Another possible factor is CPU or memory resources; insufficient CPU or memory resources could also explain the timeouts. However, investigating the specific causes of the dropped requests is beyond the scope of this paper, and further research is needed.

5 Conclusion

In this research paper, we compared the performance of a custom-made RESTful web service under various scalability options. Based on the results and analysis, we concluded that bottlenecks primarily occur in the NodeJS application rather than the MongoDB database. The most likely cause of these bottlenecks is insufficient CPU or memory resources to process each request. When many requests are sent simultaneously, some must wait for available CPU or memory resources, though no dropped requests were observed under typical conditions.

In the specific scenario involving 8,000 requests, approximately 15% were dropped due to timeout errors. This issue was likely caused by the heap memory being full, which resulted in insufficient memory to queue the requests for processing. Additional memory is needed to accommodate incoming requests.

These findings indicate that to handle a higher number of requests, we need to increase the number of NodeJS application instances (horizontal scaling) or their CPU and memory resources (vertical scaling). This approach should improve the performance of the RESTful web service.

Appendix A

The source code for the Weather-API is accessible through the following link:

https://github.com/igormkdd/weather-api

References

[1] Fielding, R.T.: Representational state transfer (rest). chapter 5 in architectural styles and the design of networkbased software architectures (2000)

[2] NodeJS: Open-source, cross-platform JavaScript runtime environment (2021). https://nodejs.org/ Accessed 2021-06-15

[3] MongoDB: Build faster, build smarter (2021). https://www.mongodb.com/ Accessed 2021-06-15

[4] Adafruit: DHT Sensor (2021). https://learn.adafruit.com/dht/ Accessed 2021-06-15

[5] Hill, M.D.: What is scalability? ACM SIGARCH Computer Architecture News **18**(4), 18–21 (1990)

[6] Gusev, M.: Scalable dew computing. Applied Sciences **12**(19), 9510 (2022)

[7] Sadalage, P.J., Fowler, M.: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional, Boston, MA (2012)

[8] Docker: Develop faster. Run anywhere. (2022). https://www.docker.com/ Accessed 2022-06-15

[9] Kubernetes: Production-Grade Container Orchestration (2022). https://kubernetes.io/ Accessed 2022-06-15

[10] Hong, X.J., Yang, H.S., Kim, Y.H.: Performance analysis of restful api and rabbitmq for microservice web application. In: 2018 International Conference on Information and Communication Technology Convergence (ICTC), pp. 257–259 (2018). IEEE

[11] Choi, M.: A performance analysis of restful open api information system. In: International Conference on Future Generation Information Technology, pp. 59–64 (2012). Springer

[12] Isha, A.S., Revathi, M.: Automated api testing. International Journal of Engineering Science **20826** (2019)

[13] The Apache software foundation: Apache JMeter (2021). https://jmeter.apache.org/ Accessed 2021-06-15

[14] Mozilla: Time to first byte (2022). https://developer.mozilla.org/en-US/docs/Glossary/time_to_first_byte/ Accessed 2022-06-15

[15] The Apdex Users Group: APDEX (2022). https://www.apdex.org/ Accessed 2022-06-15

[16] Postman: Build, test, Devug, Document, Monitor, Publish APIs togeather (2021). https://www.postman.com/ Accessed 2021-06-15

AI-Driven Approach to Educational Game Creation

Stanka Hadzhikoleva[0000-0001-5902-2624], Maria Gorgorova[0009-0001-5356-987X], Emil Hadzhikolev[0000-0001-8730-1313], George Pashev[0000-0001-8148-4737]

University of Plovdiv Paisii Hilendarski, Plovdiv, Bulgaria
[1] stankah@uni-plovdiv.bg, [2] maria.gorgorova@uni-plovdiv.bg, [3] hadjikolev@uni-plovdiv.bg, [4] georgepashev@uni-plovdiv.bg

Abstract. The use of games in the educational process has a positive impact on learner engagement and motivation, as well as on learning outcomes. With the advancement of artificial intelligence technologies and their increasing accessibility, even educators without technical skills can create educational games. This article proposes a method for using the AI chatbot ChatGPT to create educational games. The process is iterative and includes steps such as initial requirement formulation, testing, adding new functionalities or design elements, improvements, and bug fixes. Specific examples of interactive games and the corresponding cognitive skills they develop are discussed.

Keywords: educational games, creation of games, ChatGPT.

1 Introduction

With the digitalization of education, educational games are finding increasingly widespread application in e-learning. When carefully integrated into the learning process, they increase student engagement. Experiments have shown that students engage more seriously with the material when they have the opportunity to participate in game activities [1, 2]. Educational games can be used as a powerful tool to create an emotional connection between the learning content and the students [3]. The competitive element in games stimulates learners to achieve set goals, helping educators identify students who need additional support [4].

The literature presents many ideas and best practices for using educational games in various fields and subjects.

Hui and Mahmud conducted a study on the impact of games in mathematics education. Two types of cognitive domains (knowledge and mathematical skills) and five types of affective domains (achievement, attitude, motivation, interest, and engagement) were identified. The study shows that game-based learning has a positive impact on students and their mathematics results [5].

In [6], the application of game-based learning in English language courses was investigated. The results show a significant improvement in the interaction between students and the teacher, as well as in the acquisition of the material.

Educational games also find wide application in healthcare education. In [7], the results of a quantitative mapping of educational games in health education through

bibliometric analysis of publications in the scientific databases Web of Science and Scopus are presented. A clear trend of increasing popularity of educational games in health sciences education has been established.

Numerous studies show various best practices for using educational games in engineering education [8], business education [9], social science teaching [10], and others. All of this indicates the future wide application and development of technologies for the creation of educational games.

Educational games have established themselves as effective methods for increasing student motivation and engagement [11, 12]. There are numerous studies highlighting the benefits of using games in the learning process, including improved retention, critical thinking, and collaboration [13, 14]. Traditional methods for creating educational games require significant resources and technical skills, which can limit teachers' ability to use them widely [15]. With the development of AI, and particularly language models like ChatGPT, the possibilities for automating and simplifying the creation of learning materials are increasing [16, 17].

Developing various educational games aimed at learners with different levels of knowledge and different learning styles is a challenging task that requires a lot of time and specific programming skills. The rapid development of artificial intelligence technologies and their accessibility allows educators, even without technical knowledge, to create games.

The purpose of this article is to demonstrate how artificial intelligence can be used for the quick and easy creation of educational games that enhance learning. To showcase the potential of AI in creating educational materials that can be easily modified and customized to meet the needs of different groups of learners, a simple flashcard game has been chosen as an example. This is particularly useful in a learning environment where teachers' time and technical skills may be limited.

2 Creating Educational Games with ChatGPT

ChatGPT is an artificial intelligence chatbot based on a language model created by OpenAI [18]. It can communicate with a person, understand questions, the context of provided text, and respond accordingly. It is capable of analyzing and combining information from various sources to construct complete and accurate answers. ChatGPT can be used for various tasks such as translations, text generation, summaries, answering questions, generating program code, error detection, and more. ChatGPT can be utilized in education to create educational games. These games can be interactive and tailored to specific learning objectives.

In the conducted experiments, we use ChatGPT version GPT-4o. Its use is based on a subscription model, with costs varying depending on the plan chosen by the user, but there is also a free version. We use a plan priced at $20 per month, which provides access to GPT-4o, GPT-4o mini, GPT-4, unlimited prompt generation, early access to new features, access to advanced data analysis, file uploads, vision and web browsing, DALL-E image generation, creation and use of custom GPT, and more.

Creating a game is an iterative process, not a one-time act. The study employs an experimental approach to create an educational game using ChatGPT. First, we defined the key requirements for the game, including the target audience, game type, format, and structure. Then, we used multiple prompt cycles in ChatGPT to generate the game code, with each cycle adding new specifications and improvements. The final code was tested for errors to ensure functionality and compatibility.

Initially, the teacher needs to provide ChatGPT with some basic requirements, such as:

- ***Scope of the game:*** What is the educational material on which the game will be based? The teacher can provide educational content, but can also specify only a topic, in which case the chatbot will use its own knowledge base.
- ***Target group:*** What age group or knowledge level is the game intended for?
- ***Format of the game:*** Type of game – quiz, puzzle, flashcards, etc.?
- ***Number of questions/levels:*** How many questions or levels should the game include?
- ***Structure of questions/levels:*** Will the questions be multiple choice, open-ended, a combination of different types, etc.?
- ***Correct answers and explanations:*** Depending on the game, an additional requirement can be set – providing learners with the correct answers and brief explanations for them.
- ***Design:*** If there are specific design requirements, colors, style, etc.
- ***Additional game components:*** e.g., timer, points, rewards, etc.

Example command: Create an educational game on the topic of "Relational Databases" for university students! The game should be a quiz with 10 multiple-choice questions. Each question should have 4 answers, with one correct answer. Each question should have a brief explanation of the correct answer. The game should have a simple and clean design, using light colors. Include a 60-second timer for each question and a scoring system that shows the result at the end!

ChatGPT creates a web application based on HTML, CSS, and JavaScript, which runs in a web browser. The generated code should be copied into 3 text files – index.html, styles.css, and script.js. The game starts from the index.html file.

In some cases, the generated code does not work correctly. The user can request an error check with a command such as:

"This game is not working, check for errors!"

In this case, ChatGPT will regenerate the game.

The user can specify various game details in a dialog mode, for example:

"Change the game to include explanations for incorrect answers after each question!"

"Each question should be on a separate page."

"Display the number of correct answers so far on each page", etc.

It is important to note that the answers generated by ChatGPT have a fixed maximum length, which depends on the version of the language model used, the platform, the purpose of the answer, and more. In some cases, ChatGPT generates a partial response. In such situations, a command can be given to continue generating the code. This can be done, for example, with the command:

"Continue!"

From a technological perspective, creating a game with ChatGPT goes through several stages:

1. ***Request Processing.*** The chatbot receives the request text, which contains the game requirements (theme, target audience, game format, specific requirements).
2. ***Analysis and Understanding of Requirements.*** The content of the request is analyzed, and key words and concepts related to the theme are extracted.
3. ***Content Generation.*** Game elements are created depending on the type of game. This can include questions and answers, puzzles, scenarios, characters, etc.
4. ***Creation of Game Structure.*** This includes three components:
    - HTML Structure: The main structure of the web page that will contain the game is created. This includes basic elements like the title, instructions, game field, buttons, and other elements according to the game type.
    - CSS Styling: The styling of the web page is created to make it attractive and user-friendly.
    - JavaScript Logic: JavaScript code is created, containing the game logic. This includes functions for user interaction, checking the correctness of actions, managing game elements, displaying results, etc.
5. ***Content Integration.*** HTML, CSS, and JavaScript are combined.
6. ***Testing and Debugging.*** At this stage, the game is tested, and any discovered errors or bugs are corrected.
7. ***Finalization and Presentation of the Result.*** The finished game is provided to the user.

3 Types of Games and Training Higher-Order Thinking Skills

With ChatGPT, various types of educational games can be created, suitable for different educational goals and strategies, as well as learning methods. The following types of games are among the more popular:

- ***Quiz Games:*** These are interactive tests consisting of a series of questions that learners must answer. Each game usually includes multiple questions, answers, and feedback for correct and incorrect answers. The goal is to test and reinforce learners' knowledge on a specific topic.
- ***Puzzle Games:*** These are logical tasks for arranging pieces, connecting dots, solving numerical problems, etc. The goal is to stimulate logical thinking and problem-solving skills. Generally, learners solve puzzles through a "trial-and-error" mode, which encourages their persistence and patience.
- ***Flash Cards:*** These are most often used for memorization and review. Each card has two sides – one side contains a question or term, and the other provides an answer or explanation. These games can include text, images, audio, video, or other interactive content.
- ***Matching:*** In these games, learners must match pairs of elements, such as terms and their definitions or questions and answers. Different elements are

typically visualized, which need to be matched through dragging and dropping or clicking. The goal is to find the correct correspondences between elements, helping learners to reinforce their knowledge.

- ***Fill-in-the-Blanks:*** In these tasks, learners need to fill in missing words or phrases in code, sentences, or texts. The text is usually pre-prepared with blanks that need to be filled with the correct word or term. The goal is to test and improve learners' understanding of a specific topic through the contextual application of knowledge.

In Table 1, the presented types of games, example tasks in the field of programming, and the corresponding cognitive skills trained are given.

Table 1. Types of Games and Trained Cognitive Skills

Type of Game	Examples in the Field of Programming	Trained Cognitive Skills
Quiz Games	Quiz game with questions about basic concepts, terms, and definitions.	Knowledge – remembering basic terms and concepts.
	Quiz game with questions about syntax and identifying errors in given code.	Understanding – identifying and understanding syntactic errors.
	Quiz game that tests knowledge of using different algorithms and data structures in specific situations.	Application – applying knowledge to use specific algorithms and data structures in different situations.
Puzzle Games	Puzzles where players must arrange code fragments to create a working program.	Analysis – understanding the structure of a program and logically arranging the code.
	Puzzles where players must find and correct errors in given code.	Analysis – identifying and correcting errors in the code.
	Puzzles where players must use loops to solve logical tasks.	Synthesis – combining different logical approaches to solve tasks.
Flash Cards	Flash cards with basic terms and their definitions.	Knowledge – remembering terms and definitions.
	Flash cards with names of functions and methods and their purposes.	Understanding – understanding the purposes of different functions and methods.
	Flash cards with design patterns and their explanations or applications.	Application – applying design patterns in different situations.
Matching	Players match code fragments with the expected results of their execution.	Application – predicting the result of code execution.

	Players match different algorithms with their descriptions and uses.	Understanding – understanding the purpose and use of different algorithms.
	Players match terms with their definitions.	Knowledge – remembering terms and definitions.
Fill-in-the-Blanks	Players fill in missing keywords in program code.	Knowledge – remembering basic syntactic constructs.
	Players fill in missing parameters in functions and methods.	Application – applying knowledge of function and method parameters.
	Players fill in missing parts of logical expressions in conditional constructs.	Analysis – understanding and completing logical expressions.

The reviewed examples demonstrate that games can be used to develop a range of cognitive abilities.

4 Practical Example for Creating Games for Database Education

Let's consider a specific example of creating a flashcard game. The game consists of two levels: level 1 and level 2.

Level 1. The player needs to memorize 6 concepts related to relational databases, continuously clicking on the cards to view the term or the definition. This way, the player learns independently and self-tests.

Level 2. In this level, the knowledge acquired from the first level is tested – the player must match the term with its definition, with a time limit of 3 minutes. Once the player correctly matches all the concepts, they successfully complete the game.

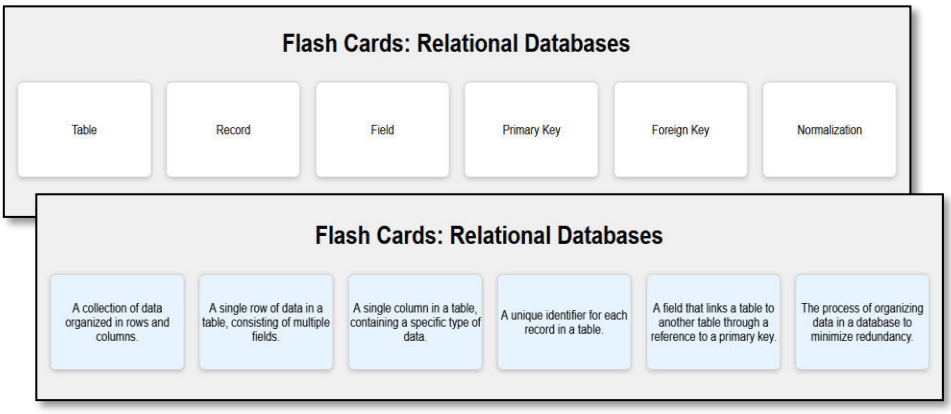


Fig. 1. First version of the game

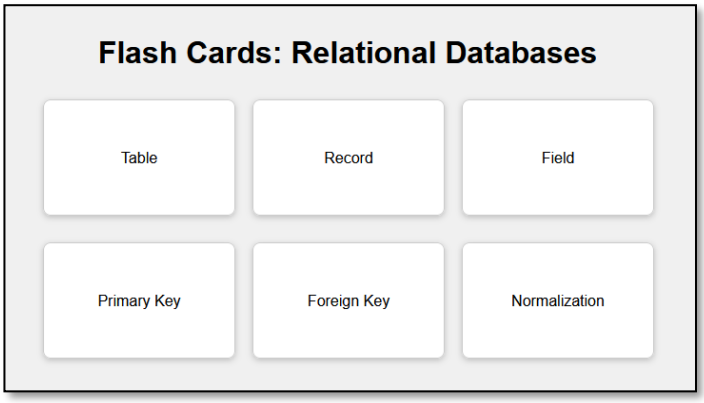


Fig. 2. Flash cards arranged in two rows

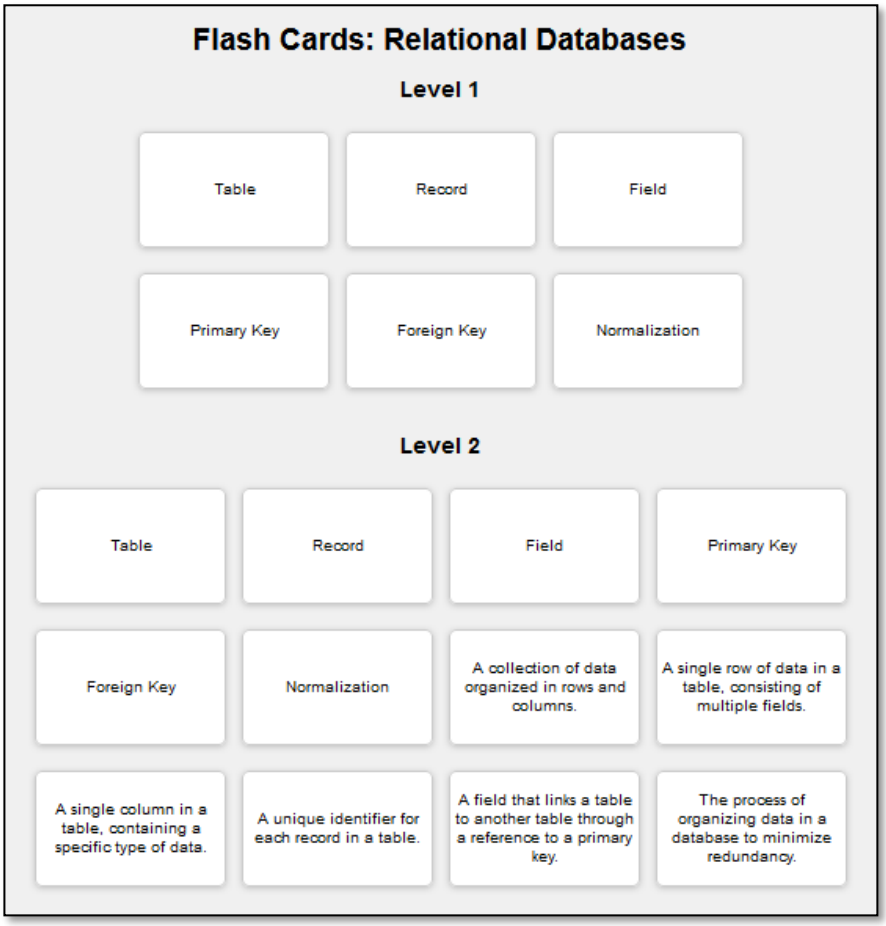


Fig. 3. Adding a new level

Request 1: Create an educational game on the topic "Relational Databases" for university students! The game should be of the "flash cards" type. Create 6 flash cards. The front of the card should have a term, and the back should have a definition. When clicking on the card, it should flip to show either the front or the back, allowing the player to continuously view the term or the definition, which helps in learning. The game should have a simple design, using light colors. Use HTML, CSS, and JavaScript to create the game! (see fig. 1)

Request 2: Arrange the cards in two rows! (see fig. 2)

Request 3: Add another level to the game! In this level, the player must match terms with definitions by clicking on the cards. There should be logic to check the match when two cards are flipped, with known correct matches between terms and their definitions. For this purpose, separate each term and each definition from level 1 into individual cards. The front of the card should have a term/definition, and the back should be just a background. Arrange the cards in 3 rows and 4 columns. The player should be allowed to flip only two cards one after the other. When there is a match between a term and its definition on two flipped cards, they disappear. If there is no match, the player flips two new cards until all matches are found. (see fig. 3)

Request 4: Separate the levels into different screens! Include a 3-minute timer for completing level 2. Flip the flash cards in level 2 so that the terms and definitions are not visible, only the background.

In request 1, the basic concept of the game is provided, and the result is shown in Fig. 1. After request 2, the flash cards are arranged in two rows (Fig. 2). Request 3 describes the concept of creating an additional level (Fig. 3). The result of request 4 is the separation of levels into different screens and the addition of buttons for navigation between levels (Fig. 4), as well as a timer for level 2. The updated two levels of the game are presented in Fig. 5 and Fig. 6.

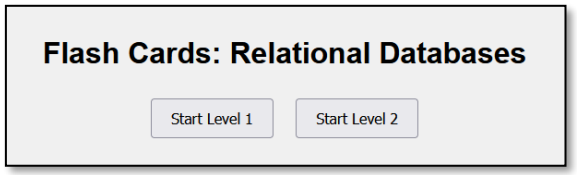


Fig. 4. Navigation buttons between levels

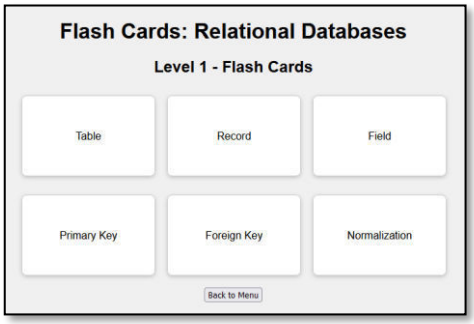


Fig. 5. Last version of the first level

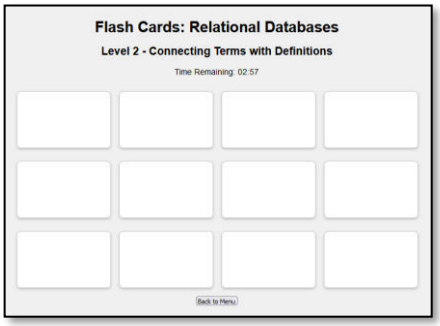


Fig. 6. Last version of the second level

In the second level, the player is allowed to flip only two cards at a time. When a match between a term and its definition is found (Fig. 7), the two cards disappear from the screen (Fig. 8). After matching all the term-definition pairs within the allotted time, the player receives a success message. If the player fails to match the pairs within the allotted time, they receive a failure message.

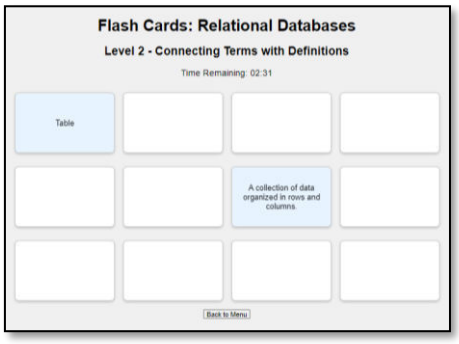


Fig. 7. Found match

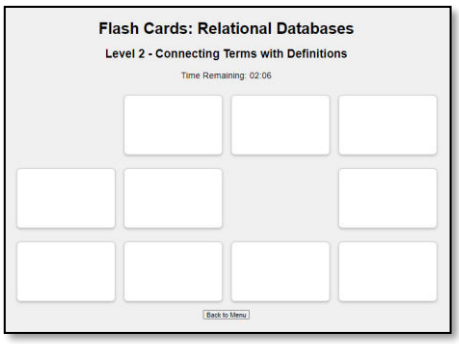


Fig. 8. Found match disappears

Thus, the provided requests were understandable for ChatGPT, and it successfully created the desired game. It is important to give the most descriptive requests possible for the best results.

In summary - the game creation process includes 4 prompt cycles to generate the complete code for the final version of the game. Each prompt cycle provides new details or clarifications to the game, allowing the model to refine the code until the desired functionality is achieved. The final code for the created web game is saved in three separate files (HTML, CSS, and JavaScript). The game was tested in Mozilla Firefox and Google Chrome web browsers. The code can be tested and used on any standard web server or local machine with a web browser installed. It can be uploaded to a publicly accessible webpage or an internal platform to be available to students.

The development of educational games using AI involves the following stages: (1) Defining the requirements (game topic, target audience, type of game); (2) Creating an initial version using an AI tool like ChatGPT; (3) Testing the game in a real environment; (4) Making corrections and improvements; (5) Finalizing and deploying in the learning environment. This process can be repeated multiple times to achieve an optimal final product. In the classical process, the development of educational games requires significantly more time and resources, including a team of programmers, designers, and educators. The AI-based process allows an individual teacher to create and customize games with minimal effort and technical knowledge, reducing development time and costs.

Flashcard games can be an effective tool for learning basic concepts and terms. They are suitable for introducing new topics. However, it is important to note that for older learners or more complex learning materials, this type of game may be less effective. To overcome these limitations, it is advisable to combine flashcards with other, more complex games that involve deeper understanding and analysis of the material.

5 Discussions

The use of ChatGPT for creating educational games offers several benefits. First, it significantly reduces development time and facilitates the creation of learning resources even by individuals without specialized technical skills. Second, the tool allows for rapid customization of games according to the needs of students and the curriculum. However, there are also some limitations. For example, the current version of ChatGPT may generate code with errors that require additional checking and corrections. Additionally, the simplicity of the generated games, such as flashcards, may not be sufficient for more complex learning objectives that require greater interactivity and critical thinking.

To ensure the effective use of AI-generated games in a learning environment, teachers should have tools for monitoring students' progress. This could include integrating game results into LMS like Moodle or Google Classroom. For example, each game could be set up to send results to the instructor in real-time, allowing for tracking the progress of individual students. Another option is for teachers to create different levels of game complexity based on students' progress. Reward and motivation systems, such as leaderboards, achievement badges, and additional challenges, can be added to maintain a high level of student engagement. This helps create a sense of competition and progress among learners.

This article focuses on demonstrating the capabilities of ChatGPT for creating educational games and presenting approaches for integrating AI into the learning environment. The main emphasis is on the practical aspects of using AI for the rapid generation of learning resources that can be adapted and customized according to the needs of students. However, we acknowledge the need for more in-depth research to evaluate user satisfaction, the functionality of the games, and their impact on the learning curve. Future work in this direction includes a comparison between AI-created and traditionally created games, as well as empirical studies to measure the effect of these games on students' cognitive development.

6 Conclusion

Using ChatGPT to create educational games offers numerous advantages for educators and their learners. Games make learning more enjoyable, stimulate motivation, and encourage active participation from learners. Educators can create various types of games, such as quizzes, puzzles, flash cards, matching games, fill-in-the-blank games, etc., which enrich the learning process and make lessons more interesting. Through these games, learners develop higher-order thinking skills.

ChatGPT automates the creation of educational games, saving educators time and allowing them to focus on teaching. This increases their productivity and the quality of education. It is significant that even people without technical skills can quickly create interactive educational games by giving natural language commands to ChatGPT. They can easily update and improve the created games based on feedback from students and

their learning outcomes. This promotes the widespread use of this technology in education.

Acknowledgments. This work was partly funded by the MUPD23-FMI-021project of the Research Fund of the University of Plovdiv "Paisii Hilendarski".

Disclosure of Interests. The authors declare no conflicts of interest.

References

1. Lam, P., Tse, A. Gamification in Everyday Classrooms: Observations From Schools in Hong Kong, Frontiers in Education, **6** (2022). https://doi.org/10.3389/feduc.2021.630666.
2. Hartt, M., Hosseini, H., Mostafapour M.: Game On: Exploring the Effectiveness of Game-based Learning, Planning Practice & Research, **35**(5) (2020).
3. Hose, D.: Game On: The Ultimate Guide to Gamify Learning Programs, Adobe eLearning (2023).
4. Chandra, S.: 14 Ways to Gamify Student Engagement & Learning, CampusGroups (2021).
5. Hui, H., Mahmud, M.: Influence of game-based learning in mathematics education on the students' cognitive and affective domain: A systematic review, **14** (2023).
6. Putri, V., Muhamad, A.: Pemanfaatan Digital Game Base Learning Dengan Media Aplikasi Kahoot.It. INSPIRASI: Jurnal Ilmu-Ilmu Sosial, **16**(2), 141–150 (2019).
7. Yıldız, M., Yildiz, M., Kayacık, A.: Rising gamification in health education: A bibliometric study, Nurse Education in Practice, **78** (2024).
8. Milosz, M., Milosz, E.: Gamification in Engineering Education – a Preliminary Literature Review, 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 2020, 1975-1979 (2020).
9. Goi, C.: Gamification in business education: Visualizing bibliometric networks analysis, Journal of Education for Business, **98**, 229-241 (2023).
10. Campillo-Ferrer, J.-M., Miralles-Martínez, P., Sánchez-Ibáñez, R.: Gamification in Higher Education: Impact on Student Motivation and the Acquisition of Social and Civic Key Competencies. Sustainability, **12**(12), (2020).
11. Cavinato, A.G., Hunter, R.A., Ott, L.S. et al. Promoting student interaction, engagement, and success in an online environment. Analytical and Bioanalytical Chemistry **413**, 1513–1520 (2021). https://doi.org/10.1007/s00216-021-03178-x
12. Hodges, L.C. Student Engagement in Active Learning Classes. In: Mintzes, J.J., Walter, E.M. (eds) Active Learning in College Science. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-33600-4_3
13. Balalle, H., Exploring student engagement in technology-based education in relation to gamification, online/distance learning, and other factors: A systematic literature review, Social Sciences & Humanities Open **9** (2024) 100870, ISSN 2590-2911, https://doi.org/10.1016/j.ssaho.2024.100870.
14. Zainuddin, Z., Chu, S. K. W., Shujahat, M., Perera, C. J. The impact of gamification on learning and instruction: A systematic review of empirical evidence. Educational Research Review **30** 100326 (2020). https://doi.org/10.1016/j.edurev.2020.100326
15. Plass, J. L., Homer, B. D., Kinzer, C. K. Foundations of game-based learning. Educational Psychologist **50**(4), 258–283 (2015). https://doi.org/10.1080/00461520.2015.1122533

16. Gill, S., Xu, M., Patros, P. et al. Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots, Internet of Things and Cyber-Physical Systems **4**, 19-23 (2024) ISSN 2667-3452. https://doi.org/10.1016/j.iotcps.2023.06.002.
17. Memarian, B., Doleck, T. ChatGPT in education: Methods, potentials, and limitations, Computers in Human Behavior: Artificial Humans **1(2)** (2023) 100022, ISSN 2949-8821. https://doi.org/10.1016/j.chbah.2023.100022.
18. ChatGPT Homepage, https://chat.openai.com/, last accessed 2024/06/30.

Session 5

Feature Selection Methods in Obesity Prediction: An Experimental Analysis

Aleksandra Sretenović, Marija Đukić [0000-0002-1136-4278], Ana Pajić Simović [0000-0002-9058-8260], Ognjen Pantelić [0000-0002-8925-4976]

University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia
as20233203@student.fon.bg.ac.rs, marija.djukic@fon.bg.ac.rs, ana.pajic.simovic@fon.bg.ac.rs, ognjen.pantelic@fon.bg.ac.rs

Abstract. This paper explores the application of machine learning in predicting obesity, a significant global health concern. We specifically examine the impact of three feature selection methods — InfoGain, Chi-squared, and ReliefF, on the performance of classification models using Random Forest and Logistic Regression algorithms. By analyzing an obesity dataset categorized into three and seven classes, we identify key features that contribute to model accuracy. The models are evaluated using several metrics: Accuracy, Precision, Recall, Specificity, Sensitivity, and Balanced Accuracy. The findings highlight the role of feature selection in model performance, with the Random Forest algorithm achieving the highest accuracy rate of 96.7%.

Keywords: feature selection, machine learning, classification algorithms, obesity

1 Introduction

The role of machine learning (ML) in healthcare continues to prove its significance and efficacy in various critical areas. By analyzing vast amounts of patient records, lab results, and treatment histories, ML uncovers patterns and trends that otherwise are easy to miss. This process aids in early disease detection, enables personalized treatment plans, improves patient outcomes, and reduces healthcare costs [25]. In this paper, we explore how ML can be used for predicting obesity.

Overweight refers to an excess of fat deposits. Obesity is a chronic and intricate condition characterized by excessive fat accumulation that can negatively impact health. Both are diagnosed by calculating body mass index (BMI) using the formula weight divided by height [28]. The transition from being lean to becoming obese triggers changes in adipose tissue, leads to chronic inflammation and increases the risk of cardiovascular diseases, and contributes to conditions such as stroke. Moreover, obesity is a major factor in insulin resistance, a key element in type 2 diabetes and metabolic syndrome. Additionally, obesity is linked to various cancers including colorectal, pan-

creatic, kidney, and endometrial [8]. According to the [28], obesity among adults worldwide has doubled since 1990, and obesity among adolescents has increased fourfold, resulting in 2.5 billion overweight adults of whom 890 million are living with obesity.

The research objective of our study is to evaluate the effectiveness of feature selection methods Chi-squared, InfoGain, and ReliefF in predicting obesity using Random Forest (RF) and Logistic Regression (LR) algorithms. The research purpose is to demonstrate that these methods can identify relevant features crucial for accurate obesity prediction, leading to high-performance models. This is achieved through experiments in Weka, assessing how feature selection methods impact model performance as dataset complexity increases from 3 to 7 classes.

This paper is structured as follows: in Section 2 we review the literature. Section 3 outlines the dataset, ML algorithms, and feature selection methods that are used. The experimental results and discussion are presented in Section 4 and Section 5 respectively, while Section 6 concludes the paper.

2 Related Work

Feature selection methods such as InfoGain, ChiSquare, and ReliefF have been applied in various domains to identify significant features, such as in cancer data [14], heart disease [21], bank data [23], and network traffic [26]. Moreover, numerous studies have employed Random Forest [5, 7, 19, 26] or Logistic Regression [4, 5, 10, 19, 21, 26] algorithms to construct predictive models, often evaluating their effectiveness through metrics like accuracy, precision, recall, and F1 score. In addition, many studies have utilized the Weka software for model development [1, 10, 14, 15, 18, 21], highlighting its role as a popular tool for machine learning tasks, including feature selection and model building.

[12] reviewed feature selection methods for medical dataset classification, highlighting challenges in balancing feature relevance and computational complexity and stressing the importance of efficient feature selection.

[7] studied obesity in Bangladesh and classified it as low, medium, or high, using 80% of their dataset for training and the rest for testing, comprising 1100 entries from diverse sources. They tested various ML algorithms and found that LR had the highest Accuracy of 97,09% after applying PCA. [19] used R software to evaluate LR and RF algorithms on an imbalanced dataset of obesity risk factors, predicting obesity as a binary classification problem. Resampling techniques were employed, with RF outperforming LR in Precision, Recall, F1 score, and Balanced Accuracy, particularly with imbalanced data.

[15] emphasized feature selection's importance in ML and proposed a new algorithm based on conditional mutual information. The testing was performed in Weka. [18] focused on breast cancer classification Accuracy using feature selection and ML algorithms. Their focus was on wrapper selection methods in Weka, noting increased Accuracy with feature selection for Bayes Network but decreased Accuracy for SVM. [21] used Weka and algorithms Bayes Net, LR, SGD, and KNN with feature selection methods Chi-squared, ReliefF, and Symmetrical uncertainty to predict if the patient has heart

disease. [26] studied Intrusion Detection Systems performance with feature selection methods Chi-squared, InfoGain, and Recursive Feature Elimination coupled with different ML classifiers. Feature selection methods improved model performance across various classifiers, confirming the importance of feature selection methods. However, the study [23] investigated how the GINI index and InfoGain affect classification Accuracy in the Decision Tree classifier algorithm and concluded that irrespective of dataset imbalance, the classification Accuracy remains consistent between models using the GINI index and InfoGain. [14] compared the PCA-IG model with traditional feature selection methods Gain Ratio, ReliefF, and CfsSubset on breast cancer data using Weka. PCA-IG outperformed in Accuracy, Precision, Recall, and training time when it comes to models built with other classifiers.

When it comes to obesity prediction, several studies have explored the effectiveness of different ML algorithms and feature selection methods. [1] explored ML algorithms for classifying childhood obesity in 6-year-old school children in Malaysia using classifiers: Naïve Bayes, Bayes Net, J48, MLP, and SMO. Feature selection methods used are CfsSubsetEvaluator and Consistency in Weka. The study found that feature selection methods with genetic search enhanced accuracy, with J48 achieving the highest Accuracy at 82,72%. Similarly, [10] analyzed obesity risk factors using PRMT in Weka, identifying Naïve Bayes with 99,2% Accuracy as the best classifier for predicting obesity risk based on factors such as age, BMI, and lifestyle. Chi-square was used to select relevant features and the classifiers built were Naïve Bayes, KNN, Kstar, ZeroR, Random Tree, and LR.

Some studies have utilized UCI Machine Learning dataset for obesity-related research. [3] proposed a model that integrates data mining techniques, including Extremely Randomized Trees, Multilayer Perceptron, and XGBoost, implemented in Python to detect and predict obesity levels. The dataset utilized originates from the UCI Machine Learning Repository. Similarly, [5] used the same UCI dataset and applied machine learning algorithms such as LR, RF, Decision Tree, SVM, Gradient Boosting, and Ada Boost. Based on evaluation metrics like accuracy, precision, recall, and F1 score, the results indicated that the Logistic Regression model achieved the highest prediction accuracy. Additionally, [4] employed Decision Trees, Logistic Regression, and KNN for prediction using the same dataset. However, studies [4] and [5] do not employ feature selection methods, whereas [3] utilizes Recursive Feature Elimination as a wrapper-type feature selection algorithm.

In comparison to prior studies, our work shares several similarities and notable differences. Similarities include the utilization of RF and LR machine learning algorithms and datasets similar to those commonly used in obesity prediction studies. Additionally, we adopt feature selection methods akin to those explored in prior works. However, our study stands out in several aspects. Firstly, we examine a dataset with 3 classes, diverging from the predominant focus on binary classification. Furthermore, we extend our investigation to a dataset with 7 classes, offering a perspective on model performance across a broader spectrum of obesity classification. Finally, our work introduces an expanded set of evaluation metrics. While previous research often relies on Accuracy, Precision, and Recall, we incorporated Sensitivity and Specificity metrics and emphasized the use of Balanced Accuracy.

3 Research Methodology

This section is divided into three parts: a description of the dataset, an overview of the machine learning algorithms used, and an explanation of the feature selection methods employed.

3.1 Dataset

The dataset was obtained from the UCI Machine Learning Repository [16], containing 2111 instances and 17 features. Feature descriptions are given in Table 1.

Table 1. Dataset feature description

Feature name	Type	Values	Description
Gender	Nominal	Male, Female	
Age	Numeric	14 – 61	
Height	Numeric	145 – 199	Height in cm
Weight	Numeric	39 – 173	Weight in kg
Family_history	Nominal	Yes, No	
FAVC	Nominal	Yes, No	Frequent caloric food intake
FCVC	Integer	1 – 3	Frequency of vegetables
NCP	Integer	1 – 4	Number of main meals
CAEC	Nominal	No, Sometimes, Frequently, Always	Food between meals
SMOKE	Nominal	Yes, No	
SCC	Nominal	Yes, No	Caloric consumption monitoring
CH2O	Numeric	1 – 3	Daily water intake
FAF	Numeric	0 – 3	Physical activity frequency
TUE	Numeric	0 – 2	Time using technology
CALC	Nominal	No, Sometimes, Frequently, Always	Consumption of alcohol
MTRANS	Nominal	Public Transport, Automobile, Walking, Bike, Motorbike	

The original target variable had seven values: Insufficient_weight, Normal_weight, Overweight_level_I, Overweight_level_II, Obesity_level_I, Obesity_level_II, and Obesity _level_III. Based on these values, a new target variable named Obesity_Risk was created, containing three possible values: Not_obese (Insufficient and Normal_weight), Overweight (Overweight_level_I and Overweight_level_II), and Obese (other values). The dataset was prepared by removing inadequate values, resulting in

1984 instances. The removed instances included unrealistic values for weight or height and values that were inconsistent when considering weight, height, and age together. Among these, there are 895 obese individuals, 546 overweight individuals, and 543 individuals with normal or insufficient weight. The inclusion of Weight and Height as features in the model was guided by recommendations from an internal medicine specialist and a comprehensive review of existing literature. Notably, studies [7, 10] have also incorporated weight in obesity prediction models.

The experiment was conducted in Weka. Based on the previous studies [10, 14], we chose Chi-squared, InfoGain, and ReliefF. These feature selection methods were employed to determine the 5, 8, and 12 most relevant features. Subsequently, classifiers were created using these features and the RF and LR algorithms. Our dataset encompassed 3 classes initially, with subsequent testing performed using the same dataset but expanded to include 7 original classes.

3.2 Machine Learning Algorithms

Random Forest and Logistic Regression are widely used machine learning algorithms for building predictive models across various domains [4, 5, 19, 21, 26], displaying strong predictive capabilities and consistent performance. These algorithms have also proven effective in prior research [7, 10] for predicting obesity. Our decision to use RF and LR was motivated by these studies, leading us to select these two algorithms as the foundation for our analysis.

A Random Forest is an ensemble learning method consisting of decision trees, collectively forming a “forest”. Each decision tree within the forest is constructed using a random subset of features at each node. During the classification phase, each tree provides a vote and the class receiving the majority of votes is selected as the final prediction [9].

Logistic regression is a statistical method for analyzing the relationship between an outcome and multiple explanatory variables. This approach calculates each variable's impact on the odds ratio of the observed event, enabling the examination of how different factors collectively influence the outcome, avoiding the pitfalls of analyzing variables in isolation [22].

3.3 Feature Selection Methods

Feature selection methods belong to a filter category that evaluates the relevance of features based on the inherent properties of the data. They are characterized by their speed, scalability, and independence from particular learning algorithms, requiring selecting features once and then assessing their effectiveness using different classifiers [12]. The selection methods used in this paper are filter methods: InfoGain, Chi-squared, and ReliefF.

To understand InfoGain, it is required to explain entropy. Defined by [20], entropy is a measure of the uncertainty or randomness in a dataset. In classification tasks, entropy quantifies the amount of impurity or disorder in a set of examples. If a dataset

contains instances that belong to different classes, the entropy will be higher. Conversely, if all instances belong to a single class, the entropy will be zero, indicating no uncertainty.

InfoGain is a metric that quantifies the reduction in entropy achieved by splitting the data based on a particular feature. It is used to determine how well a feature separates the data into classes. InfoGain is calculated as the difference between the entropy of the dataset before the split and the weighted sum of the entropies after the split. A higher InfoGain indicates that the feature is more useful for classification as it reduces uncertainty (or entropy) about the target class after the split [23]. In other words, the feature is more informative for classifying the instances.

The Chi-square statistic is a test that measures the degree of association between categorical variables. It evaluates how much the observed data deviates from what would be expected under the null hypothesis, which assumes that the two categorical variables are independent. A high value indicates a significant difference between the observed and expected frequencies, suggesting that there is a strong association between the variables. Low value, on the other hand, suggests that the observed and expected frequencies are close, indicating that the variables are likely independent or have a weak association.

In machine learning, the Chi-square test is commonly used to select the most important features when dealing with categorical data. The Chi-square statistic is computed for each feature by comparing the observed frequencies (actual data) with the expected frequencies (what would be expected if the feature was independent of the class labels). Features are ranked based on their Chi-square values. Features with higher Chi-square values are considered more relevant as they have a stronger correlation with the target variable [13]. This ranking can guide the selection of features to use in model building. Although the Chi-square test is a powerful tool for feature selection, it does have some limitations such as assuming independence of observations, sensitivity to sample size and applicability to categorical data only.

The Relief algorithm is a feature selection method that assesses how well features distinguish nearby instances. The key idea behind Relief is to evaluate the relevance of features by considering their ability to separate instances that are similar (neighbors) but belong to different classes. The algorithm randomly selects an instance from the dataset. The nearest hit is the closest instance to the selected instance that belongs to the same class. The nearest miss is the closest instance to the selected instance that belongs to a different class. For each feature, the algorithm updates a weight based on its ability to distinguish between the selected instance and its nearest hit and miss [11]. If a feature has a similar value for the selected instance and the nearest hit (same class), it is less useful and its weight is decreased. After sampling different instances, the algorithm aggregates the feature weights, ranking them according to their ability to differentiate between instances of different classes.

While the original Relief algorithm is effective, it has several limitations. ReliefF is an extension of the original algorithm designed to address its limitations. Key enhancements in ReliefF are (1) multi-class capability, (2) dealing with missing values, (3) use of multiple neighbors, (4) noise resilience and (5) weight update mechanism [17].

4 Experimental Results

For both datasets, selection methods InfoGain, Chi-squared, and ReliefF were employed using the entire dataset. This approach was chosen because if feature selection is conducted solely on the training set, the selected features or their importance rankings may vary significantly with different random states of the train-test split. This variability can lead to inconsistencies in feature selection, making it difficult to generalize the importance of features. Additionally, evaluating feature importance on the entire dataset provides a more accurate assessment of which features are generally influential. This approach is supported by several studies in literature, including [6, 15, 24].

The experiment was done in Weka, using 10-fold cross-validation, meaning that the dataset is divided into 10 equal-sized subsets. Then, the model is iteratively trained on nine of these subsets and its performance is evaluated on the remaining subset [2]. This method was chosen because it provides a more robust estimate of the model's performance compared to a single train-test split.

The metrics observed are Accuracy, Precision, Recall, Specificity, Sensitivity, and Balanced Accuracy. Accuracy is the measure of correctly classified instances. Precision reflects the accuracy of positive predictions, while Recall (Sensitivity) quantifies the model's ability to identify positive instances. Specificity evaluates the model's capability to correctly identify negative instances. Balanced Accuracy is the mean accuracy considering Sensitivity and Specificity, offering a balanced assessment of model performance, especially in scenarios with imbalanced class distributions [27]. Apart from Accuracy, Precision, and Recall which are commonly observed metrics, we incorporated the Specificity metric to assess the model's ability to differentiate instances not belonging to a specific class. This is crucial in applications where the cost of false positives is high, such as in medical diagnostics. Furthermore, we included Balanced Accuracy, which can be a valuable metric when there is not a similar balance among classes within the dataset. Balanced Accuracy provides a more equitable measure of performance by considering both Sensitivity and Specificity, ensuring that the model is not biased towards the majority class. This is essential for creating robust models that perform well across all classes.

In this section, the findings are discussed first for the 3-class dataset, followed by an analysis of the results obtained from the dataset with 7 classes.

4.1 Results with 3-class Target Variable

Feature ranking based on the method is given in Table 2. It can be noted that Chi-squared and InfoGain give similar rankings, while ReliefF results differ. This difference can be featured in the underlying methodologies used. Chi-square and InfoGain both rely on statistical measures to assess the relevance of features. On the other hand, ReliefF focuses on evaluating feature relevance by considering values between nearest neighbors from the same and different classes.

Table 2. Feature ranking for 3-class dataset

Rank	InfoGain	Chi-squared	ReliefF
1	Weight	Weight	Weight
2	Family_history	CAEC	Family_history
3	CAEC	Family_history	CAEC
4	Age	Age	FCVC
5	NCP	NCP	NCP
6	FAF	FAF	TUE
7	FAVC	FAVC	FAF
8	TUE	TUE	Height
9	FCVC	FCVC	Age
10	SCC	Height	CH2O
11	Height	MTRANS	MTRANS
12	MTRANS	SCC	FAVC
13	CALC	CALC	Gender
14	Gender	Gender	SCC
15	CH2O	CH2O	CALC
16	SMOKE	SMOKE	SMOKE

Weight, Family_history, CAEC, and NCP consistently rank in the top five features across all three methods, meaning these features are highly influential in predicting and understanding obesity in the dataset. Age, FAF, and TUE consistently rank high, but with slight variations in their specific orders across methods. Gender, CALC, and SMOKE consistently rank towards the bottom, suggesting that they have minimal direct impact. MTRANS, FCVC, CH2O, SCC, Height, and FAVC are features whose rankings fluctuate the most across different feature selection methods. This fluctuation in rankings indicates that these features may have varying degrees of influence on predicting or understanding obesity depending on the specific methodology used for feature selection.

The ReliefF feature selection method produces distinct rankings compared to InfoGain and Chi-squared, particularly in the middle and lower ranks. While Weight, Family_history, CAEC, and NCP remain consistently influential, appearing in the top five across all methods, the ReliefF method shows variability with other features. Notably, FCVC and TUE, ranked fourth and sixth by ReliefF, contrast with their more consistent middle rankings by InfoGain and Chi-squared. Features such as Height, FAVC, and SCC exhibit significant ranking shifts under ReliefF, suggesting their influence on predicting obesity varies notably with this method.

Results for models built with InfoGain and Chi-squared feature rankings are given in Table 3. In terms of performance metrics, RF constantly outperforms LR. For 5 and 8 features, RF significantly outperforms LR, indicating that even with fewer features RF can effectively capture the complexities of the dataset better than LR. Both classifiers exhibit improvement as the number of features increases and achieve the best results for 12 features, with RF achieving the highest Accuracy at 96,6%. The Balanced Accuracy values are high, suggesting a well-rounded performance across all classes,

with minimal bias towards any specific class. Models built using Chi-squared feature ranking show the same results. The similarity in performance can be attributed to the small variation in feature rankings between Chi-squared and InfoGain, proving their correlation.

Table 3. RF and LR results using InfoGain and Chi-squared ranking

InfoGain, Chi-squared	Accuracy	Precision	Recall	Specificity	Sensitivity	BA
5 features						
RF	90,4%	90,4%	90,4%	95,4%	90,4%	92,9%
LR	83,6%	84,2%	83,6%	91,9%	83,6%	87,8%
8 features						
RF	91,6%	91,6%	91,6%	96%	91,6%	93,8%
LR	85%	85,6%	85%	92,6%	85%	88,8%
12 features						
RF	96,6%	96,6%	96,6%	98,5%	96,6%	97,6%
LR	96,2%	96,2%	96,2%	98,3%	96,2%	97,3%

Results for models built using ReliefF ranking are given in Table 4. Models demonstrate overall higher performance, but also more variations compared to results obtained using InfoGain and ChiSquare feature ranking. RF still achieves higher results than LR. For both algorithms increasing the number of features from 5 to 8 leads to significant improvements in model performance. Both models give the best results with 8 features used, with RF achieving the highest Accuracy at 96,7%. Eight features that demonstrate the best performance are: Weight, Family_history, CAEC, FCVC, NCP, TUE, FAF, and Height. The Specificity metric values, exceeding 92%, indicate the model's proficiency in minimizing false positives, while the high Balanced Accuracy suggests the model's ability to make precise predictions across all classes.

Table 4. RF and LR results using ReliefF ranking

ReliefF	Accuracy	Precision	Recall	Specificity	Sensitivity	BA
5 features						
RF	86,9%	86,9%	86,9%	93,3%	86,9%	90,1%
LR	83,8%	84,4%	83,8%	92,1%	83,8%	88%
8 features						
RF	96,7%	96,7%	96,7%	98,5%	96,7%	97,6%
LR	96,6%	96,6%	96,6%	98,5%	96,6%	97,6%
12 features						
RF	96,6%	96,6%	96,6%	98,5%	96,6%	97,6%
LR	95,8%	95,8%	95,8%	98,2%	95,8%	97%

4.2 Results with 7-class Target Variable

In this section, we show the model performance variations when using a 7-class dataset. The evaluation metrics used are Accuracy and Recall, as we primarily focus on feature selection rather than assessing overall model performance. Feature ranking based on the method is given in Table 5. Again, Chi-squared and InfoGain give similar rankings, while ReliefF results differ.

Table 5. Features ranking for 7-class dataset

Rank	InfoGain	Chi-squared	ReliefF
1	Weight	Weight	Gender
2	Age	Age	Weight
3	FCVC	FCVC	FCVC
4	Gender	CAEC	Family_history
5	CAEC	Gender	CAEC
6	Family_history	Family_history	CALC
7	NCP	Height	MTRANS
8	Height	NCP	NCP
9	CALC	FAF	Height
10	FAF	CALC	TUE
11	MTRANS	MTRANS	Age
12	TUE	TUE	FAF
13	FAVC	FAVC	FAVC
14	SCC	SCC	CH2O
15	CH2O	CH2O	SCC
16	SMOKE	SMOKE	SMOKE

Models constructed using InfoGain and Chi-squared consistently produce similar results, despite differences in feature ranking. Notably, when selecting subsets of 5, 8, and 12 features, both methods identify the same groups of features. RF still achieves higher results than LR. For both algorithms, increasing the number of features from 5 to 8 leads to significant improvements, with RF achieving the highest Accuracy at 94% with 8 features. However, Accuracy decreases going from 8 to 12 features for both classifiers (Table 6).

Table 6. RF and LR results using InfoGain and Chi-squared ranking, 7-class dataset

InfoGain, Chi-squared	Accuracy	Recall
5 features		
RF	85,4%	85,4%
LR	73,8%	73,8%

8 features		
RF	94%	94%
LR	91,5%	91,5%
12 features		
RF	93,6%	93,6%
LR	91,3%	91,3%

Using ReliefF for feature selection, there is a more pronounced improvement in Accuracy and Recall as the number of features increases (Table 7). Both RF and LR models exhibit significant enhancements in performance from 5 to 8 features and further improvements at 12 features. The best Accuracy at 93,6% is achieved with RF, using 12 features.

Table 7. RF and LR results using ReliefF ranking, 7-class dataset

ReliefF	Accuracy	Recall
5 features		
RF	79%	79%
LR	73,9%	73,9%
8 features		
RF	86,5%	86,5%
LR	75,1%	75,1%
12 features		
RF	93,6%	93,6%
LR	91,3%	91,3%

5 Discussion

In both datasets, Weight, Age, CAEC, and Family_history maintain relatively high rankings across different methods, indicating their importance. Features FCVC, Gender, MTRANS, CALC, and Height become significantly more important in the 7-class dataset compared to the dataset with 3 classes, while TUE, FAF, and FAVC lose relevance. Features SMOKE and CH2O consistently rank towards the bottom, suggesting that they have a minimal direct impact, regardless of the dataset. RF consistently outperforms LR. For the 3-class dataset, the highest accuracy of 96.7% is achieved using the ReliefF method with RF and 8 features. In the case of the 7-class dataset, RF achieves the highest accuracy of 94% using 8 features with InfoGain or Chi-squared rankings.

In a 3-class dataset, the InfoGain and Chi-squared feature selection methods demonstrate a trend of increasing accuracy as more features are added, with the best results achieved at 12 features. However, the ReliefF method exhibits a different pattern: classifiers reach their peak performance with just 8 features. Adding more features beyond this point does not improve performance, suggesting that a more streamlined feature subset may be more effective for model optimization. In contrast, in a 7-class dataset,

the patterns shift. When using InfoGain and Chi-squared for feature selection, classifiers achieve the highest accuracy with 8 attributes. However, classifiers built using ReliefF continue to improve as more features are added, peaking at 12 features.

This variation highlights how the performance of feature selection methods is influenced by the number of classes in the dataset. While InfoGain and Chi-squared methods show optimal performance at 12 features in a 3-class dataset, they perform best with 8 features in a 7-class dataset. Conversely, ReliefF, which peaks at 8 features in a 3-class dataset, reaches its highest accuracy with 12 features in a 7-class dataset. These results show that both the choice of feature selection method and the optimal number of features are closely tied to the dataset's class structure.

The study [7] presents relevant findings for comparison with our work.

Table 8. Comparison with study in [7]

ML algorithm	Research	Accuracy	Precision	Recall
LR	Our study	96,6%	96,6%	96,6%
	[7]	97,09%	97%	97%
RF	Our study	96,7%	96,7%	96,7%
	[7]	72,3%	57%	72%

The authors of [7] worked with a dataset comprising 1100 entries and 28 features to classify obesity into 3 classes: low, medium, or high. They employed ML algorithms KNN, SVM, LR, Naïve Bayes, RF, Decision Tree, Ada Boosting, MLP, and Gradient Boosting. Comparing overall performance, their study achieved a slightly higher Accuracy rate of 97,09% employing LR and PCA. In contrast, our study involves a larger dataset with 1984 entries and 16 features. We used feature selection methods instead of PCA. These methods focus on selecting a subset of features based on their relevance, whereas PCA transforms the entire set of features into new variables. Comparing algorithm performance, the authors of [7] achieved better results using LR. Our work demonstrated significantly better performance with RF, particularly in terms of Accuracy and Precision. Notably, we incorporated Balanced Accuracy as an evaluation metric, and our study achieved a rate of 97,6% using RF and 8 features selected with ReliefF.

5.1 Potential Limitations

Despite our thorough analysis, there are several potential limitations to this research that should be acknowledged. One concern is the specificity of the dataset, as the findings may not generalize well to other datasets with different characteristics. We did not include validation using datasets from other sources, which would demonstrate the generalizability and robustness of the results beyond the dataset used in this study. Furthermore, the focus on the effectiveness of Chi-squared, InfoGain, and ReliefF for feature selection, without including results from models that do not use these methods, may limit understanding of the full impact of feature selection on model performance. The

study could also benefit from the inclusion of additional feature selection methods and machine learning models, which we plan to explore in future research.

6 Conclusion

Our study demonstrated that feature selection methods can effectively reduce the number of model features while maintaining comparable performance in classification models, especially in the context of obesity prediction. Through experimentation in Weka, we have identified several key features, including Weight, Age, FCVC, CAEC, and Family_history. Notable findings highlight the superiority of RF over LR and the best model built with RF having an Accuracy rate of 96,7%. The transition from a 3-class to a 7-class dataset emphasizes the increased significance of the feature selection method for the Accuracy metric. InfoGain and Chi-square methods maintain consistent and reliable feature rankings and groupings, showcasing their suitability for this purpose. ReliefF exhibited variability in feature rankings compared to InfoGain and Chi-squared, contributing to noticeable improvements in model performance as the number of selected features increased. This variability highlights ReliefF's efficacy in identifying relevant features that contribute to enhancing model accuracy and robustness. Effective feature selection greatly enhances model performance overall.

In the future, our research aims to expand into data preprocessing techniques to enhance input data quality, address class imbalance issues, and incorporate additional feature selection methods. We will continue to refine our models and explore their applicability to more extensive datasets. Additionally, some more advanced tools could be utilized for creating and testing models, which would allow for further analysis on how these tools could enhance the results and insights of the study.

Acknowledgments. This paper is funded by the Faculty of Organizational Sciences, University of Belgrade.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdullah, F. S., Manan, N. S., Ahmad, A., Wafa, S. W., Shahril, M. R., Zulaily, N., . . . Ahmed, A. (2016). Data Mining Techniques for Classification of Childhood Obesity Among Year 6 School Children. *Recent Advances on Soft Computing and Data Mining* (pp. 465–474). Bandung: Springer. doi:https://doi.org/10.1007/978-3-319-51281-5_47
2. Berrar, D. (2019). Cross-Validation. In *Encyclopedia of Bioinformatics and Computational Biology* (Vol. 1, pp. 542-545). Elsevier. doi:10.1016/B978-0-12-809633-8.20349-X
3. Choudhuri, A. (2022, January). A hybrid machine learning model for estimation of obesity levels. In International Conference on Data Management, Analytics & Innovation (pp. 315-329). Singapore: Springer Nature Singapore.

4. Cui, T., Chen, Y., Wang, J., Deng, H., & Huang, Y. (2021, May). Estimation of Obesity levels based on Decision trees. In 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM) (pp. 160-165). IEEE.
5. Devi, K. N., Krishnamoorthy, N., Jayanthi, P., Karthi, S., Karthik, T., & Kiranbharath, K. (2022, January). Machine Learning Based Adult Obesity Prediction. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.
6. Elemam, T., & Elshrkawey, M. (2022). A highly discriminative hybrid feature selection algorithm for cancer diagnosis. The Scientific World Journal, 2022(1), 1056490.
7. Ferdowsy, F., Rahi, K. A., Ismail , J., & Habib, T. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences, 2*. doi:https://doi.org/10.1016/j.crbeha.2021.100053
8. Fruh, S. M. (2017). Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. *Journal of the American Association of Nurse Practitioners*, S3-S14. doi:10.1002/2327-6924.12510
9. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier Inc. doi:https://doi.org/10.1016/C2009-0-61819-5
10. Hossain, R., Mahmud, H., Hossin, A., Noori, S., & Jahan, H. (2018). PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques. *International Conference on Computational Intelligence and Data Science Proceedings*, *132*, pp. 1068-1076. doi:https://doi.org/10.1016/j.procs.2018.05.022
11. Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and new algorithm. *AAAI-92: Tenth National Conference on Artificial Intelligence Proceedings* (pp. 129-134). San Jose: AAAI. doi:https://dl.acm.org/doi/10.5555/1867135.1867155
12. Mwadulo, M. W. (2016). A Review on Feature Selection Methods For Classification Tasks. *International Journal of Computer Applications Technology and Research, 5*(6), 395-402. doi:10.7753/IJCATR0506.1013
13. Na, S., An, X., Yan, C., & Ji, S. (2020). Incremental Feature Reduction Method Based on Chi-Square Statistics and Information Entropy. *IEEE Access, 8*, 98234-98243. doi:10.1109/ACCESS.2020.2997013
14. Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature Selection for Classification using Principal Component Analysis and Information Gain. *Expert Systems with Applications, 174*. doi:https://doi.org/10.1016/j.eswa.2021.114765
15. Phyu, T. Z., & Oo, N. N. (2016). Performance Comparison of Feature Selection Methods. *MATEC Web of Conferences.* doi:10.1051/matecconf/20164206002
16. Repository, U. (2019). *Estimation of Obesity Levels Based On Eating Habits and Physical Condition.* Retrieved April 2024, from UC Irvine Machine Learning Repository: https://doi.org/10.1016/j.dib.2019.104344
17. Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning, 53*, 23-69. doi:https://doi.org/10.1023/A:1025667309714
18. Saoud, H., Ghadi, A., Mohamed, G., & Abdelhakim, B. A. (2019). Using Feature Selection Techniques to Improve the Accuracy of Breast Cancer Classification. *The Proceedings of the Third International Conference on Smart City Applications*, (pp. 307-315). doi:10.1007/978-3-030-11196-0_28
19. Sewpaul, R., Awe, O. O., Dogbey, D. M., Sekgala, M. D., & Dukhi, N. (2023). Classification of Obesity among South African Female Adolescents: Comparative Analysis of Logistic Regression and Random Forest Algorithms. *International Journal of Environmental Research and Public Health*, *21*(1), 2.
20. Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*(3), 379-423. doi:https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

21. Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. (2020). Exploring feature selection and classification methods for predicting heart disease. *Digital Health*. doi:10.1177/2055207620914777
22. Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12-18. doi:10.11613/BM.2014.003
23. Tangirala, S. (2020). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm. *(IJACSA) International Journal of Advanced Computer Science and Applications, 11*(2), 612-619.
24. Tariq, M. A. (2024). A Study on Comparative Analysis of Feature Selection Algorithms for Students Grades Prediction. Journal of Information and Organizational Sciences, 48(1). https://doi.org/10.31341/jios.48.1.7
25. Tekieh, M. H., & Raahemi, B. (2015). Importance of Data Mining in Healthcare: A Survey. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1057-1062). Paris: IEEE. doi:http://dx.doi.org/10.1145/2808797.2809367
26. Thakkar, A., & Lohiya, R. (2021). Attack classification using feature selection techniques: a comparative study. *Journal of Ambient Intelligence and Humanized Computing*, *12*(1), 1249-1266.
27. Vujovic, Z. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications Volume, 12*(6), 599-606. doi:10.14569/IJACSA.2021.0120670
28. WHO. (2024, March 1). *Obesity and overweight*. Retrieved April 14, 2024, from World Health Organization: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

A New AFIB Detection Method with Deep Neural Networks

Sara Gjorgjieva[1,2], Ana Angjelevska[1,2], Dimitar Trajanov[2][0000−0002−3105−6010], Marjan Gusev[1,2][0000−0003−0351−9783]

[1] Innovation Dooel, Skopje, North Macedonia
{sara.gjorgjieva, ana.angjelevska}@innovation.com.mk
[2] Sts Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, North Macedonia
{dimitar.trajanov, marjan.gushev}@finki.ukim.mk

Abstract. Early detection of atrial fibrillation (AFIB), characterized by irregular heart rhythms, can potentially save thousands of lives annually [20]. Numerous studies have attempted to address this challenge using various algorithms and methodologies. This paper introduces a novel and robust method that enhances the precision of AFIB detection. Neural Networks (NN) are widely recognized for their ability to identify complex electrocardiogram (ECG) patterns, facilitating the accurate detection of irregular heart rhythms. This study explores multiple conventional machine learning algorithms alongside neural network approaches to develop a robust AFIB detection system. To evaluate the effectiveness and robustness of the proposed method, the MIT BIH arrhythmia database (MITDB) and the Long-term Atrial Fibrillation Database (LTAFDB), both available on the Physionet site. Our approach presents a novel feature engineering technique, employing specific features such as a time series of the differences in successive heart rates (dBPM) rather than traditional inter-beat intervals. The model analyzes inputs labeled into five distinct classes, using only clean intervals identified through a new labeling method for training while treating others as outliers. This strategy ensures a more accurate and reliable AFIB detection process. The best-performing model, a Deep Neural Network (DNN), achieved an F1 score of 95.18% for binary atrial fibrillation classification on a benchmark dataset distinct from the training set. Models trained using our approach have demonstrated superior performance than previously published models.

Keywords: atrial fibrillation, arrhythmia, deep learning, ECG

1 Introduction

Atrial fibrillation (AFIB) is a type of cardiac arrhythmia that occurs in the heart's atria due to disrupting the normal cycle of the heart's electrical impulses. It typically arises from increased automatism in a group of cells within the atria, causing their contractions to exceed the regular rate of impulses per minute and resulting in complete electrical

disorganization within the atria. These abnormal impulses override the heart's natural pacemaker, disrupting its ability to maintain a regular rhythm and leading to an irregular heart rate [2].
AFIB prevents the heart muscle from relaxing properly between contractions, reducing its efficiency and overall performance. The condition manifests as a highly irregular pulse rate. It can lead to numerous health issues, including rapid heartbeat, weakness, and dizziness, leading to blood clots, stroke, heart failure, and other heart-related complications [2].
Detecting AFIB requires the ability to identify specific ECG characteristics. AFIB is an irregular rhythm of heartbeats that seems normal in width but varies in height and spacing without preceding polarization P-waves. Most state-of-the-art studies focus on detecting an irregular rhythm in the ECG without considering the absence of P-waves, which are very difficult to detect due to their very low amplitude. An irregular rhythm may have patterns such as those classified with Bigeminy (B) and Trigeminy (T), while AFIB is an irregular rhythm without patterns. State-of-the-art analysis shows that most Neural Network approaches use RR intervals as the time between consecutive heartbeats. In this paper, we use a strategy based on differences in heart rates between consecutive heartbeats as a novel feature proven to outperform other approaches. We used multiple techniques to achieve a robust real-time medical device application, including cross-validation and an additional validation dataset to prevent overfitting. For this purpose, we used the standard *MIT-BIH arrhythmia ECG benchmark database (MITDB)* [16] and the *Long-Term Atrial Fibrillation ECG Database (LTAFDB)* [18] available on the Physionet site [11].
The paper follows the next structure. Related work in Section 2 analyses the similar approaches. Section 3 presents the methods in this research, explaining the datasets, feature engineering, machine learning models, and evaluation methodology. Section 4 addresses the performance evaluation from the experiments. Discussion in Section 5 compares the achieved results with the related work. Section 6 addresses the conclusion and gives directions for future work.

2 Related Work

Search for an effective algorithm to detect AFIB results in numerous approaches. Our analysis shows the application of different signal processing and Machine Learning (ML) models, including NNs. Evaluation methods differ from using proprietary datasets via standards, such as EC57 [4] or EIC 60601-2-47 [12] to derive meaningful evaluation metrics. Authors who use the same dataset for training and testing reveal high results, although these models are not robust enough. Applying the same model to a different patient performs very poorly. An example includes [15] declaring an F1 score of 98.97% with their NN model. A model created by combining a recurrent NN and long-short-term memory (LSTM) [10] was trained and tested on the same MITDB dataset with a 10-fold cross-validation to reach an F1 score of 98.35%.

To overcome the robustness and achieve applicability, the researchers used the patient-wise method, with totally different datasets for the training and testing. For example, [26] uses different combinations of convolutional NNs and long-short-term memory (LSTM) frameworks. Although using a 70/30 splitting criteria on the same MITDB dataset, this approach reveals an F1 score of 82%. Recurrent NNs for detecting AFIB [23] on the same dataset (MITDB) obtain an F1 score of 83.00%. A distributed deep NNs [7] was used on a proprietary dataset collecting records from multiple hospitals and wearable ECG devices to detect AFIB with a 5-fold cross-validation, achieving an F1 score of 90.70%.

Our approach belongs to the class of patient-wise data split for training and testing datasets. Contrary to other NN solutions, our approach uses features based on differences in rates from consecutive heartbeats, while the others use intervals between consecutive heartbeats.

We plan three experiments: the first using the same dataset for training and testing, the second using separate ECG benchmark tests, and the third dividing the MITDB into two equal datasets. Many other papers typically use cross-validation splits.

3 Methods

3.1 Datasets

Databases in this research are fundamental in studying heart diseases and are widely referenced in the scientific community, publicly available from the Physionet website [1].

MITDB ECG database [16] contains 30-minute ECG recordings from 48 patients, totaling 109,483 annotated heartbeats. This research excludes four records from paced patients (102, 104, 107, and 217). Consequently, the total number of annotated heartbeats analyzed is 97,924. It is a golden benchmark for evaluating newly developed algorithms and ML models, and in this research, training the new model uses MITDB. Only 10% of records consist of AFIB rhythm episodes.

LTAFDB database [18] consists of 24 to 25-hour ECG recordings from 83 individual patients diagnosed with AFIB, with over half of the records with AFIB rhythm. This database contains 8,903,169 annotated heartbeats, excluding two paced patients (6 and 113); we excluded them, totaling with 8,682,368, but because it is enormous compared to the other dataset, we used a shortened version where we included 943,953 annotated heartbeats.

The records from both databases contain predominantly Normal Sinus Rhythm (NSR), including episodes of Bigeminy (B), Trigeminy (T), Bradycardia (SBR), Tachycardia (SVT or SVTA), and a small portion of other atrial and ventricular arrhythmia. This research excludes episodes of Ventricular Fibrillation (VFIB) and Ventricular Flutter (VFL) from the analysis. Both databases are imbalanced, which is a common challenge in medical data. Addressing this imbalance will be crucial in training and evaluating our models effectively for atrial fibrillation detection.

3.2 Feature Engineering

Feature engineering is crucial in developing machine learning models, especially for detecting complex conditions such as atrial fibrillation (AFIB) from heart rate data. Feature engineering aims to extract features from raw data that can effectively capture the underlying patterns and differences between normal heart rhythms and AFIB episodes. For AFIB detection, various features derived from electrocardiogram (ECG) data enhance the model's ability to distinguish between different cardiac rhythms.

This research focuses on features derived from the intervals between heartbeats to develop robust machine-learning models for AFIB detection. These features include RR intervals and additional measures of irregularity and complexity.

- *RR Intervals* represent the time difference between two consecutive heartbeats. It is a fundamental feature in most AFIB detection research because RR interval variations indicate irregular heart rhythms. Analyzing the sequence of RR intervals helps identify the erratic patterns typical of AFIB [21].
- *GZIP Compression Length* transforms each bucket formed from the time series value into a UTF-8 character and concatenates them into a string of 41 characters. The length of the compressed string serves as a feature, providing a measure of the sequence's complexity and variability.
- *Shannon Entropy (ShEn)* quantifies the unpredictability or randomness in the buckets formed from time series values. It is calculated as (1), where n_i is the number of values in the i-th bucket, and N is the total number of intervals. Higher entropy indicates more significant irregularity, which is characteristic of AFIB episodes.

$$ShEn = -\sum_{i=0}^{N-1}(\frac{n_i}{N}) * log(\frac{n_i}{N}) \tag{1}$$

- *Sample Entropy (SaEn)* is calculated using the TSFEL library processing the buckets formed from time series as input. Sample Entropy measures the complexity of time-series data, with higher values suggesting more irregular and complex patterns, typical of AFIB.

By leveraging these features, our approach aims to effectively capture the nuances of different heart rhythms, enhancing the model's ability to distinguish between AFIB and NonAFIB episodes. This comprehensive feature engineering strategy is pivotal in improving the accuracy and reliability of AFIB detection using machine learning techniques.

To train the model effectively, we use only clean segments [9] and apply a majority rule, where intervals with more than 51% of heartbeats labeled as 1 (indicative of AFIB) are classified as AFIB.

3.3 Ensemble Machine Learning Models

Ensemble Machine Learning Models combine the predictions from multiple models to improve accuracy and robustness. In our earlier work [25],

we analyzed several ensemble ML models, which were also employed in this study for AFIB detection.
To ensure the robustness and effectiveness of these models, we employed a 10-fold cross-validation technique along with hyperparameter tuning.

Decision Tree (DT) creates a model based on simple decision rules inferred from the data features. It recursively splits the data into subsets based on feature values, maximizing the separation between classes at each node. Decision Trees are intuitive and easy to interpret, making them a good baseline for initial model development in AFIB detection [6]. The following hyperparameters were tuned in the optimization process: maximum depth in range (1,20), minimum samples per leaf (2,6), random state [42], criterion [gini, entropy], maximal features [sqrt,log2, None], and CCP alpha [0.0, 0.01 0.1, 0.2].

Random Forest (RF) is an ensemble method that builds a collection of Decision Trees, each trained on a randomly selected subset of the training data. By aggregating the predictions from multiple trees, Random Forest reduces overfitting and improves generalization. This method is particularly effective in dealing with noisy data and capturing complex heart rate variability patterns indicative of AFIB. [13]. The same hyperparameter settings were used as in DT, maximum depth in range (1,20), minimum samples per leaf (2,6), random state [42], criterion [gini,entropy], maximal features [sqrt,log2, None], and CCP alpha [0.0, 0.01 0.1, 0.2]

XGBoost (XGB) stands for Extreme Gradient Boosting as a powerful ensemble learning technique that builds decision trees in parallel, optimizing each tree based on the errors of the previous trees. XGBoost incorporates regularization techniques to prevent overfitting, making it highly effective for large datasets and complex feature spaces. Its ability to compute individual tree errors concurrently enhances model performance and generalization, which is crucial for accurately detecting AFIB episodes amidst varying heart rhythms [8]. The hyperparameter tuning optimized the learning rate in the range [0.01, 0.1, 0.16, 0.2], number of estimators [100, 200, 300], max depth range(1,20), colsample bytree [0.5, 0.7 0.9], min child weight [1,2], and regularization parameter reg alpha [0.01, 0.1, 1].

AdaBoost (AB) stands for Adaptive Boosting that combines multiple weak learners sequentially. It assigns higher weights to misclassified instances, forcing subsequent models to focus more on these challenging cases. This iterative process improves the model's ability to correct errors and enhances its robustness. It supports the classification of imbalanced datasets, such as those with fewer AFIB episodes compared to NonAFIB episodes, by focusing learning on the minority class [14]. We optimized the number of estimators in the range [100, 200, 300] and learning rate [0.01, 0.1, 0.2, 0.3] within the hyperparameter tuning.

CatBoost (CB) means Categorical Boosting that handles categorical features effectively through an ordered boosting approach. It builds decision trees by considering the order of observations and incorporating categorical data directly into the model. CatBoost is known for its training efficiency and ability to handle high-cardinality categorical features, making it suitable for complex datasets encountered in AFIB detection. We selected the loss function [Logloss] and evaluation metric [AUC]. The hyperparameter tuning optimized the learning rate [0.01, 0.1, 0.2] and colsample by level [0.5, 0.7, 0.9].

3.4 Deep Learning Model Development

We also employed deep learning (DL) models to detect atrial fibrillation (AFIB) from ECG data.

Recurrent Neural Network (RNN) are designed to process sequential data where the order of inputs matters, such as time series data. In AFIB detection, RNNs are advantageous because they can learn temporal dependencies and relationships between heartbeats, capturing the irregular patterns characteristic of AFIB episodes. RNNs maintain a hidden state updated at each time step, allowing them to retain information across long sequences [22]. They can effectively handle sequential data, such as ECG signals [3] because it is designed to maintain temporal dependencies between inputs.
The RNN model used in this study consists of several layers. The SimpleRNN Layer is the initial layer in the network that processes input data as a sequence of time steps, capturing temporal patterns. The *dense layers* are created with Leaky ReLU Activations. After the SimpleRNN layer, the model incorporates six fully connected (Dense) layers. These layers use leaky ReLU (Rectified Linear Unit) activation functions to introduce non-linearity, allowing the network to learn more complex data patterns. The *Dropout Layers* are interspersed between the dense layers to mitigate overfitting. These layers randomly deactivate some neurons during training, forcing the network to learn more robust features. *L2 Regularization* [19] is applied to the dense layers to penalize large weights, further controlling overfitting by smoothing the model. Learning Rate Scheduler is a dynamic scheduler that adjusts the learning rate during training epochs to enhance model convergence and performance. *Optimizer and Loss Function* is realized by the Adam optimizer for efficient gradient-based optimization and binary cross-entropy loss function to measure the error in binary classification tasks (AFIB vs. NonAFIB). The *training parameters* include multiple epochs with a batch size of 128. The validation set, derived from a separate test set, assesses the model's generalization performance during training.

Siamese Neural Network (SNN) is based on a specialized architecture that learns to measure similarity or dissimilarity between two input instances, regardless of the specific task. In the context of AFIB

detection, SNNs can be trained to compare segments of ECG data and determine whether they represent similar rhythm patterns (e.g., AFIB vs. NonAFIB) [17].
The *twin network architecture* in the SNN consists of two identical branches, each processing ECG signals. The architecture is designed such that both branches have the same structure and shared weights, ensuring that they learn the same features. Each branch separately consists of three fully connected Dense layers and four more merged layers. The *feature extraction and comparison* is done during training, where both branches receive pairs of ECG signals (one set from AFIB episodes and another from NSR episodes). The outputs from both branches are combined and used for the final classification. Yje *loss Function and optimization* uses the binary cross-entropy loss function, suitable for binary classification tasks. The Adam optimizer updates the network weights during training, minimizing classification errors. The *key advantage* is the dual-branch SNN architecture that allows the model to effectively learn temporal patterns and similarities between pairs of inputs, which enhances its ability to differentiate between AFIB and NSR episodes.

Deep Neural Network (DNN) is designed to learn and represent complex data patterns through multiple layers of neurons. It is particularly effective at extracting hierarchical features from raw input data, essential for making accurate predictions or classifications in AFIB detection. By leveraging deep architectures, DNNs can capture high-level abstractions and fine-grained details from ECG data, enhancing the model's ability to distinguish between different heart rhythms [24].
We implement DNN architecture using Keras to extract and process features from the input data through multiple layers. The *dense layers with leaky ReLU activations* are crucial for capturing non-linear relationships in the input data and distinguishing between normal sinus rhythm and AFIB episodes. The —em dropout layers are applied to the dense layers to prevent overfitting. This regularization technique helps the model generalize unseen data better by randomly dropping units during training. The *Optimizer and Loss Function* is realized by the Adam optimizer for training and minimizes the binary cross-entropy loss function. This setup is well-suited for binary classification tasks and helps the model learn effectively. The *training strategy* optimizes the model's ability to classify AFIB episodes accurately.

Hyperparemeter optimization includes the following ranges: dropout rate [0.1 - 0.5], optimizer [adam,sgd], epochs [5-50], batch size [32 - 256], activation [relu, leaky relu] and layers config [3,9,12,18].

3.5 Evaluation Methodology

Evaluating the performance of machine learning models in detecting atrial fibrillation (AFIB) is critical, especially given the inherent class

imbalance in AFIB datasets, where NonAFIB episodes often outnumber AFIB episodes. We use the F1 score to assess model performance, a metric particularly suited for imbalanced classification problems.

The **F1 score** is calculated as the harmonic mean of sensitivity (SEN) and precision, providing a balanced measure that considers both false positives and false negatives. This is especially important in medical diagnoses, where missing an AFIB episode (false negative) or incorrectly identifying a NonAFIB episode as AFIB (false positive) can have significant clinical implications. The F1 score is calculated by (2), Where: **Sensitivity** (SEN), also known as the True Positive Rate (TPR), measures the proportion of actual positive cases (AFIB episodes) correctly identified by the model calculated by (3), and **Precision**, also known as Positive Predictive Value (PPV), measures the proportion of predicted positive cases versus all positives, calculated by (3).

$$F_1 = \frac{2 * PPV * SEN}{PPV + SEN} \tag{2}$$

$$SEN = \frac{TP}{TP + FN} \qquad PPV = \frac{TP}{TP + FP} \tag{3}$$

Duration-Based Evaluation Method [11] evaluates the duration of correctly identified AFIB episodes instead of the heartbeat-oriented evaluation, providing a more nuanced understanding of the model's ability to detect AFIB over varying time intervals.

Each analyzed interval is assigned a value, and a specific post-processing procedure is applied to smooth out short sequences with alternating labels. This smoothing process helps mitigate the impact of noisy predictions and more accurately delineates the beginning and end of AFIB episodes. By refining these sequences, the model's detection of AFIB episodes aligns more closely with clinically meaningful patterns, providing a more reliable evaluation of its real-world applicability.

This dual approach ensures that the evaluation captures the model's ability to correctly identify AFIB cases and its effectiveness in recognizing AFIB episodes amidst varying heart rhythms.

3.6 Experiments

We specify three experiments based on different training/validation/testing datasets.

Experiment 1 uses a random selection of data samples and forms three subsets by an 80/10/10 split ratio. This splitting method lacks robustness and generality since it includes data from the same patient in all three datasets, so the model will be evaluated with data for the patient for which it was trained and validated. Experiment 1a uses MITDB and Experiment 1b LTAFDB.

Experiment 2 uses different datasets where the first dataset forms the training and validation dataset with a ratio of 80/20 split and a different dataset for testing. We use MITDB as the first dataset divided into training and validation datasets. LTAFDB is used as a testing dataset. This method ensures testing with a completely independent dataset, providing a robust measure of its performance in real-world scenarios and generalization capability across different datasets and patient populations. Experiment 2a uses MITDB for training and LTAFDB for testing, while Experiment 2b uses the opposite LTAFDB for training and MITDB for testing.

Experiment 3 uses the De Chazal Method, where the MITDB dataset is divided with the interpatient method into DS1 and DS2 datasets, specifically designed to handle the class imbalance present in the MITDB database, such that DS1 is used for training and DS2 is used for testing. The key aspect of this method is that patients included in the training dataset are not included in the testing set, preventing data leakage and overfitting. This split method is highly effective in balancing patient distribution. The DS1 dataset includes patients 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, and 230, and DS2 the remaining from the MITDB.

4 Results

Tables 1 through 5 present the results for each evaluated model across different experiments.

Table 1. Results obtained from Experiment 1a, training with MITDB, and evaluating with MITDB

Method	SEN (%)	PPV (%)	F1 (%)	Mean (%)	STD (%)
DT	91.14	87.71	89.39	73.02	32.66
RF	91.66	89.36	**90.49**	88.52	1.14
XGB	91.35	88.52	89.91	88.89	0.60
AB	93.10	86.84	89.86	88.92	0.62
CB	91.56	89.17	90.35	89.53	0.13
RNN	91.80	87.24	89.46	89.46	0.00
SNN	97.44	76.62	85.79	85.79	0.00
DNN	88.21	89.03	88.62	88.62	0.00

Among all the models, the **Deep Neural Network (DNN)** model, as shown in Table 3, which was trained on the MITDB dataset and tested on the LTAFDB dataset, stands out as the most effective. It achieved an impressive F1 score of 95.18% in detecting AFIB with separate training and testing datasets.

Table 2. Results obtained from Experiment 1b, training with LTAFDB, and evaluating with LTAFDB

Method	SEN (%)	PPV (%)	F1 (%)	Mean (%)	STD (%)
DT	96.56	96.95	96.75	96.55	0.13
RF	96.55	97.05	96.80	96.68	0.14
XGB	96.44	97.11	96.77	96.74	0.09
AB	96.82	96.69	96.75	96.68	0.06
CB	96.58	97.06	**96.82**	96.84	0.01
RNN	96.29	97.18	96.73	96.73	0.00
SNN	96.61	96.71	96.66	96.66	0.00
DNN	96.44	97.06	96.75	96.75	0.00

Table 3. Results obtained from Experiment 2a, training with MITDB, and evaluating with LTAFDB

Method	SEN (%)	PPV (%)	F1 (%)	Mean (%)	STD (%)
DT	90.90	97.01	93.85	72.99	32.65
RF	89.88	97.25	93.42	88.48	1.09
XGB	89.52	97.29	93.24	88.78	0.50
AB	91.78	97.17	94.40	88.85	0.57
CB	89.58	97.21	93.24	89.49	0.14
RNN	93.75	96.48	95.09	95.09	0.00
SNN	95.25	94.58	94.91	94.91	0.00
DNN	94.87	95.49	**95.18**	95.18	0.00

Table 4. Results obtained from Experiment 2b, training with LTAFDB, and evaluating with MITDB

Method	SEN (%)	PPV (%)	F1 (%)	Mean (%)	STD (%)
DT	96.43	58.60	72.89	96.52	0.13
RF	96.17	59.18	73.27	96.65	0.14
XGB	96.43	58.92	73.14	96.73	0.10
AB	96.92	57.46	72.15	96.65	0.06
CB	96.23	58.71	72.93	96.82	0.01
RNN	97.05	58.00	72.61	72.61	0.00
SNN	93.78	64.36	**76.34**	76.34	0.00
DNN	97.69	56.00	71.18	71.18	0.00

The **DNN** is the best-performing model due to its exceptional ability to generalize across different datasets. This model performed well on distinct datasets and demonstrated robustness by including cross-validation and an additional validation dataset in the evaluation process.

5 Discussion

Based on the results obtained from our experiments and the reviewed studies, NNs have proven effective in detecting AFIB.

Table 5. Results obtained from Experiment 3, training with DS1, and evaluating with DS2

Method	SEN (%)	PPV (%)	F1 (%)	Mean (%)	STD (%)
DT	87.35	79.93	83.47	53.10	37.58
RF	89.20	79.74	**84.21**	59.59	35.96
XGB	83.24	81.20	82.21	80.62	1.88
AB	81.29	82.29	81.79	82.22	0.97
CB	79.67	82.23	80.93	80.51	0.59
RNN	91.04	75.30	82.43	82.43	0.00
SNN	92.14	68.87	78.82	78.82	0.00
DNN	88.08	73.34	80.04	80.04	0.00

Table 6. Comparison with other papers

Related Work	Method	F1(%)	Split	Dataset
Martis et al.[15]	ANN	98.97	-	MITDB, AFDB
Wang et al.[26]	CNN+LSTM	82.00	70/30	MITDB
Teijeiro et al.[23]	RNN	83.00	-	PhysioNet/CinC 2017
Cai et al.[7]	DDNN	90.70	5 fold CV	proprietary data
Faust et al.[10]	RNN + LSTM	98.35	10 fold CV	AFDB
Artis et al.[5]	ANN	92.60	-	MITDB
Our method	RF	90.49	80/10/10+10 fold CV	MITDB
Our method	CB	96.82	80/10/10+10 fold CV	LTAFDB
Our method	DNN	95.18	separate datasets	MITDB, LTAFDB
Our method	SNN	76.34	separate datasets	LTAFDB, MITDB
Our method	RF	84.21	deChazal split	MITDB

To ensure robustness, we used different variations of training and testing of the models to prove their efficiency and accuracy. Due to the imbalance in our datasets, we selected the F1 score as our performance metric, as it is well-suited for assessing model performance on skewed datasets. We also used 10-fold cross-validation on our ensembling methods, mean and standard deviation of the classifier performance, and additional hyperparameter tuning on these methods.

The success of our approach also hinges on selecting appropriate parameters and designing an optimal NN structure tailored for AFIB detection. While the SNN, RNN, and DNN methods also performed well, the CB model demonstrated superior performance in our study, with an F1 score of 96.82%. Table 6 compares the results from related research achieved by other authors. Note that our methods did not achieve the highest scores compared to the others. Still, we aimed to attain maximal robustness using a 10-fold cross-validation and addition validation dataset. We also included hyperparameter tuning to reach the best results from our models. Additionally, we applied extra techniques to prevent overfitting.

Table 7 represents the optimal hyperparameters used to achieve the highest result for each train/test combination we used, as shown in Table 6.

Table 7. Optimal hyperparameters for the results achieved in Table 6

Model	Training	Testing	Optimal Parameters
RF	MITDB	MITDB	ccp alpha:[0.0], criterion:[gini], max depth:[12], max features:[sqrt], estimators:[300], random state [42]
CB	LTAFDB	LTAFDB	colsample bylevel:[0.5], evaluation metric:[AUC], learning rate:[0.1], loss function:[Logloss]
DNN	MITDB	LTAFDB	layers_config=[3, 9, 12, 18], dropout rate:[0.1], optimizer:[adam], epochs:[19], batch size:[128]
SNN	LTAFDB	MITDB	layers_config=[3, 9, 12, 18], epochs:[8], batch size:[128], optimizer:[adam]
RF	DS1	DS2	ccp alpha:[0.0], criterion:[entropy], max depth:[2], max features:[sqrt], estimators:[200], random state:[42]

6 Conclusion

We presented a new method for developing ML models based on newly derived features, such as Gzip length, Shannon Entropy, and Spectral Entropy, besides the series of differences in heart rates between consecutive heartbeats. Special labeling techniques divide the dataset into five classes, and the AFIB binary classification determines the start and end of an AFIB rhythm applied to the duration-based evaluation. We used the MITDB and LTAFDB datasets for the experiments and tried all the combinations to see which gave us the best results. Specifically, the decision to train the model only with clean intervals improved the model's performance. The results outperform other research compared with the same evaluation methodology.

This research shows that DNN models have great potential in biomedical signal processing, especially for detecting AFIB. Advanced deep learning techniques are promising technology after carefully designing the underlying problem. These advancements offer hope for improving medical outcomes and patient care in heart health. Refining NN methods further could lead to even more progress in AFib detection and other vital areas of medical research. However, besides the secret ingredients in developing the model, we conclude that feature engineering improved the whole process, especially selecting the training subset from the training dataset by applying a specific labeling method.

References

1. Databases. https://www.physionet.org/about/database/ (2024), accessed: 2024-04-26
2. American Heart Association: What is Atrial Fibrillation? (2024), last accessed online on 01.07.2024 at: https://www.heart.org/en/health-topics/atrial-fibrillation/what-is-atrial-fibrillation-afib-or-af
3. Andersen, R.S., Peimankar, A., Puthusserypady, S.: A deep learning approach for real-time detection of atrial fibrillation. Expert Systems with Applications **115**, 465–473 (2019)

4. ANSI/AAMI: ANSI/AAMI EC57:2012 (R2020) testing and reporting performance results of cardiac rhythm and st segment measurement algorithms (2012), last accessed online on 19.01.2024 at: https://webstore.ansi.org/standards/aami/ansiaamiec572012r2020
5. Artis, S.G., Mark, R., Moody, G.: Detection of atrial fibrillation using artificial neural networks. Master's thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and ... (1991)
6. Bin, G., Shao, M., Bin, G., Huang, J., Zheng, D., Wu, S.: Detection of atrial fibrillation using decision tree ensemble. In: 2017 Computing in Cardiology (CinC). pp. 1–4. IEEE (2017)
7. Cai, W., Chen, Y., Guo, J., Han, B., Shi, Y., Ji, L., Wang, J., Zhang, G., Luo, J.: Accurate detection of atrial fibrillation from 12-lead ecg using deep neural network. Computers in biology and medicine **116**, 103378 (2020)
8. Chen, Y., Wang, X., Jung, Y., Abedi, V., Zand, R., Bikak, M., Adibuzzaman, M.: Classification of short single-lead electrocardiograms (ecgs) for atrial fibrillation detection using piecewise linear spline and xgboost. Physiological measurement **39**(10), 104006 (2018)
9. Dojchinovski, D., Gusev, M.: Segment labeling method for ml-based afib detection (2020)
10. Faust, O., Shenfield, A., Kareem, M., San, T.R., Fujita, H., Acharya, U.R.: Automated detection of atrial fibrillation using long short-term memory network with rr interval signals. Computers in biology and medicine **102**, 327–335 (2018)
11. Gusev, M., Boshkovska, M.: Performance evaluation of atrial fibrillation detection. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 342–347. IEEE (2019)
12. IEC: IEC 60601-2-47:2012 medical electrical equipment - part 2-47: Particular requirements for the basic safety and essential performance of ambulatory electrocardiographic systems (2012), last accessed online on 19.01.2024 at: https://webstore.ansi.org/standards/aami/ansiaamiiec60601472012r2016
13. Kennedy, A., Finlay, D.D., Guldenring, D., Bond, R.R., Moran, K., McLaughlin, J.: Automated detection of atrial fibrillation using rr intervals and multivariate-based classification. Journal of electrocardiology **49**(6), 871–876 (2016)
14. Mahmood, I.S., Abdelrahman, I.A.M.: A comparison between different classifiers for diagnoses of atrial fibrillation. In: 2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). pp. 1–6. IEEE (2019)
15. Martis, R.J., Acharya, U.R., Prasad, H., Chua, C.K., Lim, C.M., Suri, J.S.: Application of higher order statistics for atrial arrhythmia classification. Biomedical signal processing and control **8**(6), 888–900 (2013)
16. Moody, G.B., Mark, R.G.: The impact of the mit-bih arrhythmia database. IEEE engineering in medicine and biology magazine **20**(3), 45–50 (2001)

17. Ng, Y., Liao, M.T., Chen, T.L., Lee, C.K., Chou, C.Y., Wang, W.: Few-shot transfer learning for personalized atrial fibrillation detection using patient-based siamese network with single-lead ecg records. Artificial Intelligence in Medicine **144**, 102644 (2023)
18. Petrutiu, S., Sahakian, A.V., Swiryn, S.: Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. Europace **9**(7), 466–470 (2007)
19. Phaisangittisagul, E.: An analysis of the regularization between l2 and dropout in single hidden layer neural network. In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS). pp. 174–179. IEEE (2016)
20. Sankari, Z., Adeli, H.: Heartsaver: A mobile cardiac monitoring system for auto-detection of atrial fibrillation, myocardial infarction, and atrio-ventricular block. Computers in biology and medicine **41**(4), 211–220 (2011)
21. SK, S.R., Kolekar, M.H., Martis, R.J.: A deep learning approach for detecting atrial fibrillation using rr intervals of ecg. Frontiers in Biomedical Technologies **11**(2), 255–264 (2024)
22. Sujadevi, V., Soman, K., Vinayakumar, R.: Real-time detection of atrial fibrillation from short time single lead ecg traces using recurrent neural networks. In: Intelligent systems technologies and applications. pp. 212–221. Springer (2018)
23. Teijeiro, T., García, C.A., Castro, D., Félix, P.: Arrhythmia classification from the abductive interpretation of short single-lead ecg records. In: 2017 Computing in cardiology (cinc). pp. 1–4. IEEE (2017)
24. Tran, L., Li, Y., Nocera, L., Shahabi, C., Xiong, L.: Multifusionnet: atrial fibrillation detection with deep neural networks. AMIA Summits on Translational Science Proceedings **2020**, 654 (2020)
25. Tudjarski, S., Ignjatov, T., Gusev, M.: A new ml-based afib detector. In: 2021 29th Telecommunications Forum (TELFOR). pp. 1–4 (2021). https://doi.org/10.1109/TELFOR52709.2021.9653409
26. Wang, J., Li, W.: Atrial fibrillation detection and ecg classification based on cnn-bilstm. arXiv preprint arXiv:2011.06187 (2020)

Session 6

Social Media Use by Businesses in Europe

Aneta Velkoska[0009-0008-2314-9795] and Atanas Hristov1[1111-2222-3333-4444]

[1] University for Information Science and Technology "St. Paul the Apostle"
Ohrid, Republic of North Macedonia
aneta.velkoska@uist.edu.mk, atanas.hristov@uist.edu.mk

Abstract. In the dynamic landscape of modern commerce, traditional advertising strategies have proven insufficient in meeting the evolving demands of global markets. Over the past decade, social media has emerged as a transformative force, reshaping business interactions and strategies worldwide. The exponential growth of social media platforms reflects broader trends in digital transformation, wherein businesses harness data-driven strategies to optimize marketing efforts, enhance customer relationships, and drive innovation. From small enterprises seizing newfound marketing avenues to large corporations solidifying their market presence, social media has become indispensable in crafting compelling brand narratives and fostering customer loyalty. This paper explores the profound impact of social media in Europe as a pivotal tool for businesses, emphasizing its unparalleled potential for outreach, networking, and interactive engagement. By regression analysis, we proved that social media usage for any purpose for any size class of enterprise follows a linear trend with a similar slope. Mixed-Effects Model analysis shows that there are significant differences in the web adoption average value respectively to the size class of enterprise and leading are the large businesses, but the differences in all types of businesses are consistent across the years. Also, this Mixed-Effects Model indicates that the usage trends between the size class of enterprise vary significantly depending on the different social media such as Social Networks, Blogs/Microblogs, and Multimedia-sharing platforms, but the upward trend is consistent across all of them and large businesses are leading the way, particularly in the use of social networks and multimedia-sharing platforms.

Keywords: Social media, Businesses, Digital transformation.

1 Introduction

We currently live in a world in which traditional advertising strategies aren't enough. Social media has gained significant momentum as a business tool in the past decade. Not only does it allow for tremendous outreach and networking, but also allows for interactivity that can be very beneficial to businesses for a variety of reasons. In the current modern societies, social media are frequently used to connect people from all around the world using the World Wide Web and the Internet, [1]. Whether it is through social network platforms like Facebook, Instagram, Twitter, LinkedIn, YouTube,

Pinterest, or various blogs, forums, or other media-sharing websites, people can now have conversations online, also called interactive dialogues, with anybody and on any subject permitting them to share their experiences and valuable information. Thanks to the personal autonomy and freedom that the Internet offers, people are actively connecting and talking about their experiences, sharing their opinions about products and services they have tested or even just heard about, [2].

The exponential growth of social media over the past decade is one of the most formidable developments in the history of commerce. Social media can also be defined as a connection between people in which they share, create, and exchange ideas on networks with an interest in relative information. The social media have made communication easy in today's world with easier ways of finding jobs online, linking business to business together, virtual meetings, online interviews, chatting, and sharing documents, pictures, videos, and other types of media, [3]. It has been an essential tool in job creation and boosting the economy of many countries. In 2011 it was revealed that when companies focus on their business and interact with individuals, they will be able to produce large quantities of "exhaust data", for example, the data that will be combined with other activities to create a by-product. Billions of individuals around the world are contributing wondrous amounts of data to social media through many devices like desktop computers, smartphones, tablets, and so on. Social Media Marketing has offered a large variety of new opportunities for companies to promote their brand, products, and services.

By looking at the enormous amount of social media campaigns, forums, e-commerce websites, blogs, and sales emails, it shows that companies of all sizes have been translating their marketing approaches to the Internet because of its accessibility to their target audience and the amount of financials required to do so, [4].

The importance of social media in business is growing at a rapid speed. With more and more people joining social media sites and using them regularly and efficiently, the social media industry is bound to become bigger in the upcoming years. As technology progresses, the social media growth is not slowing down. With such amazing growth, every business today should take advantage and leverage proper social media channels in the best possible way because their target audience is hanging around the popular social networks and they are engaging with their favorite brands and connecting with them on different levels. By connecting your business to social media, not only you generate more business but you also connect your customers better and serve them on a higher level, [5].

The primary objective of this paper is to analyze and present the usage of social media by businesses in Europe, specifically focusing on different business sizes and sectors. The study aims to provide insights into how small, medium, and large businesses in the EU 27 countries are adopting and utilizing social media platforms for various purposes. This excludes sectors such as agriculture, forestry and fishing, mining and quarrying, and the financial sector.

The data that is used for this survey is based on the results of the Eurostat Statistics, [6] from a Community survey on "ICT usage and e-commerce in enterprises". Statistics are obtained from enterprise surveys conducted by National Statistical Authorities between 2018 and 2023. The statistical observation unit is the enterprise. The sectors

covered are manufacturing, electricity, gas and steam, water supply, construction, wholesale and retail trades, repair of motor vehicles and motorcycles, transportation, and storage, accommodation and food service activities, information and communication, real estate, professional, scientific and technical activities, administrative and support activities and repair of computers and communication equipment. Enterprises are broken down by size; small (10-49), medium (50-249), and large enterprises (250 or more persons employed).

The paper is organized as follows. The next section explores the trends in social media adoption by businesses of different sizes in the EU countries from 2013 to 2023 by analyzing how small, medium, and large enterprises have integrated social media into their operations and the varying growth patterns across these business sizes.

Section 3 analyzes the trends in website adoption and functionalities among small, medium, and large businesses in the EU countries from 2018 to 2023. It aims to highlight the growth patterns and strategic importance of websites for different business sizes as part of their digital transformation efforts.

In Section 4 we identify the evolving trends in social media usage among businesses of varying sizes across EU 27 countries from 2019 to 2023, focusing on specific platforms such as social networks, enterprise blogs, microblogs, and multimedia content-sharing websites. Meaning, how small, medium, and large businesses are integrating social media into their digital strategies to enhance marketing, customer engagement, and overall business operations.

Section 5 concludes the paper.

2 Social media use by purpose and size class of enterprise in the EU countries

2.1 Usage of social media for any purpose and size class of enterprise

Over the past decade, the use of social media by businesses in the EU 27 countries has seen significant changes. Fig.1 explores the trends in social media adoption from 2013 to 2023 across small, medium, and large businesses, excluding sectors such as agriculture, forestry and fishing, and mining and quarrying, and without considering the financial sector.

A linear regression analysis on the provided data is performed, by treating the year as the independent variable (X) and the percentages for small, medium, and large businesses as the dependent variables (Y).

Small Businesses

In 2013, only a quarter (25.4%) of small businesses utilized social media for any purpose. However, by 2015, this figure rose to 32.6%, marking a 7.2 percentage point increase. This initial growth suggests that small businesses began to recognize the importance of social media as a valuable tool for marketing and customer engagement. By 2017, the adoption rate further accelerated, reaching 42.5%. This 9.9% increase

signifies a period where small businesses increasingly integrated social media into their business strategies. The steady growth continued, with 47.9% of small businesses using social media in 2019, indicating a persistent upward trend.

The most substantial growth occurred between 2019 and 2023, with the usage rate jumping to 57.1%. This 9.2 percentage point increase reflects an accelerated adoption rate, possibly driven by the broader digital transformation trends and the impact of the COVID-19 pandemic, which highlighted the necessity of digital engagement.

By regression analysis, we get the following results: the coefficient of determination r^2 = 0.973, which indicates that 97.3% of the variance in the small business is explained by the year, the F-statistic: 107.0, *p*-value = 0.00193 is below 0.05, meaning the linear relationships is statistically significant, with a slope of the linear regression line 3.20 indicating that for each year increase the percentage in the small businesses using social media increases by approximately 3.2%.

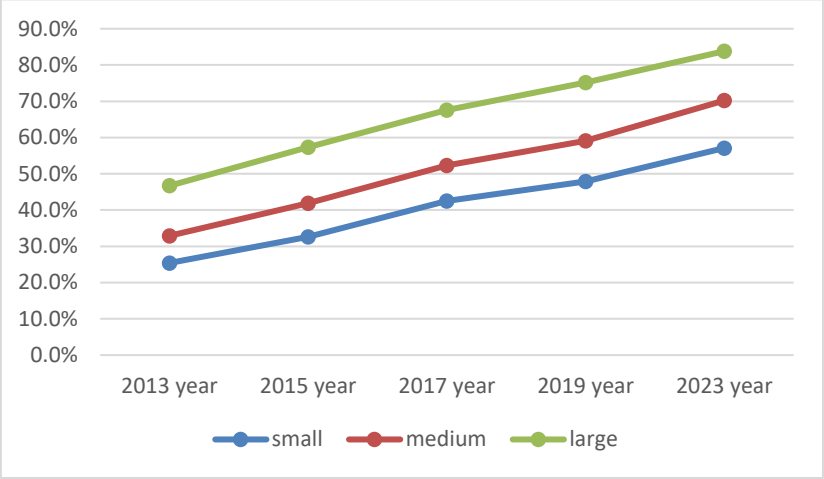


Fig. 1. Social media use by any purpose and different size class of enterprise in the EU countries, Eurostat [7]

Medium Businesses

Medium-sized businesses followed a similar upward trajectory. In 2013, 32.9% of these businesses were on social media. By 2015, the rate increased to 41.9%, a 9 percentage point rise, mirroring the initial growth pattern seen in small businesses.

The adoption rate surged to 52.3% by 2017, reflecting a significant 10.4 percentage point increase. Medium businesses, often having more resources than small businesses, could more effectively leverage social media tools and strategies during this period.

From 2017 to 2019, the usage grew to 59.1%, marking a 6.8 percentage point rise. This steady growth phase demonstrates the ongoing integration of social media into business operations. By 2023, the adoption rate soared to 70.2%, an 11.1 percentage point increase, indicating that medium-sized businesses are rapidly embracing social media as a critical component of their digital strategy.

By regression analysis, we get the following results: the coefficient of determination r^2 = 0.969, which indicates that 96.9% of the variance in the medium business is explained by the year, the F-statistic: 95.7, *p*-value = 0.00250 is below 0.05, meaning the linear relationships is statistically significant, with a slope of the linear regression line 3.76 indicating that for each year increase the percentage in the medium businesses using social media increases by approximately 3.76%.

Large Businesses

Large businesses had the highest initial social media usage rates in 2013, with 46.7% using these platforms. By 2015, this figure grew to 57.3%, a 10.6 percentage point increase, showcasing an early and robust adoption phase. The growth continued steadily, with 67.6% of large businesses on social media by 2017, marking another 10.3 percentage point rise. This period likely saw large businesses refining their social media strategies and increasing their investments in digital marketing. Between 2017 and 2019 the usage of social media increased by 7.6%, and by 2023 reached 83.8%.

By regression analysis, we get the following results: the coefficient of determination r^2 = 0.956, which indicates that 95.6% of the variance in the medium business is explained by the year, the F-statistic: 64.55, *p*-value = 0.00403 is below 0.05, meaning the linear relationships is statistically significant, with a slope of the linear regression line 3.71 indicating that for each year increase the percentage in the medium businesses using social media increases by approximately 3.71%.

The steady increase in social media usage across all business sizes under-scores the critical role of digital transformation in the EU 27 countries. r^2-squared values for all three categories are high, indicating that the year is a strong predictor of the percentage in each category, the p-values for all F-statistics are below 0.05, meaning the linear relationships are statistically significant, and the slopes for each businesses category suggest that the percentages of social media usage for small, medium, and large businesses all increase over time, with a similar rate of increase.

3 Website and functionalities by size class of enterprise in EU countries

3.1 Website Adoption and functionalities by size class of enterprise

The evolution of website adoption among businesses in the EU 27 countries from 2018 to 2023 reflects a broader trend towards digital transformation across different business sizes. We will explore how small, medium, and large businesses have embraced websites as a fundamental part of their operations, excluding sectors such as agriculture, forestry and fishing, and mining and quarrying, and not including the financial sector, see Fig. 2.

Small Businesses

In 2018, 73.8% of small businesses had established an online presence through a website. This figure saw a slight increase to 74.2% in both 2019 and 2020, indicating a steady but slow adoption rate. By 2021, the percentage of small businesses with a website rose to 75.4% and continued to increase slightly to 75.6% in 2023.

This gradual upward trend signifies that small businesses are increasingly recognizing the importance of having a website. Despite limited resources compared to larger enterprises, small businesses are investing in their digital presence to reach a broader audience, improve customer engagement, and compete more effectively in the digital marketplace. The slow but steady growth also suggests that while many small businesses are adopting websites, there may still be barriers such as cost, technical expertise, or perceived need that some businesses are overcoming at a gradual pace.

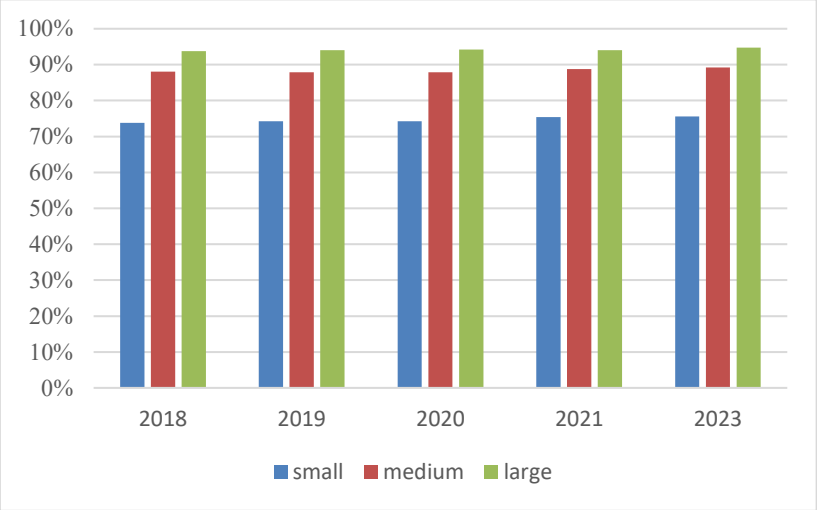


Fig. 2. Website Adoption and functionalities by size class of enterprise in the EU countries, Eurostat, [8]

Medium Businesses

Medium-sized businesses have consistently maintained higher rates of website adoption compared to small businesses. In 2018, 88.1% of medium businesses had a website. This figure experienced a slight dip to 87.9% in both 2019 and 2020 but rebounded to 88.8% in 2021. By 2023, the adoption rate had increased to 89.2%.

The high and relatively stable adoption rates among medium businesses highlight the critical role that websites play in their operations. Medium-sized businesses often have more resources than small businesses, allowing them to invest in comprehensive digital strategies that include maintaining a robust online presence. The slight fluctuations observed between 2018 and 2021 may reflect temporary market conditions or shifts in business priorities, but the overall trend indicates a strong commitment to

leveraging websites for business growth, customer engagement, and competitive advantage.

Large Businesses

Large businesses have shown the highest levels of website adoption throughout the period from 2018 to 2023. In 2018, 93.8% of large businesses had a website, with this figure increasing slightly to 94.0% in 2019 and 94.2% in 2020. In 2021, the percentage remained stable at 94.0%, and by 2023, it had risen to 94.7%.

The consistently high adoption rates among large businesses underscore the integral role of websites in their operations. For large enterprises, having a website is not just about online presence; it's a critical platform for a wide range of activities including marketing, sales, customer service, and stakeholder engagement. The near-universal adoption rate indicates that large businesses view their websites as essential tools for maintaining their market position, driving business operations, and engaging with a diverse array of stakeholders.

To account for both fixed effects, i.e. differences in website adoption between small, medium, and large businesses, and random effects, meaning year-to-year variability we use the Mixed Model analysis.

The intercept associated with large businesses is 94.14 and it represents the baseline value. The medium businesses coefficient is -5.76, meaning that the web adoption average value among medium businesses is 5.76 % lower than the large businesses, and this difference is statistically significant since $p < 0.001$. The small businesses coefficient is -19.50, so the web adoption average value among small businesses is 19.50% lower than the large category, and this difference is statistically significant since $p < 0.001$. The group variance is 0.254, so the variance due to the year effect is relatively small, suggesting that the year-to-year variability is minimal compared to the fixed effect of the type of businesses, indicating that the differences in all types of businesses are consistent across the years.

4 Social Media Usage for Specific Purposes in EU 27 Countries by Business Size

Over the years 2019 to 2023, businesses of various sizes in the EU 27 countries have shown significant changes in their social media usage. We will explore how small, medium, and large businesses have adopted social media for specific purposes in the EU 27 countries across the years 2019, 2021, and 2023. The sectors considered exclude agriculture, forestry and fishing, and mining and quarrying, and do not include the financial sector. The purposes analyzed include the use of social networks, enterprise blogs or microblogs, multimedia content-sharing websites, and the overall usage of any social media. Fig. 3 gives a graphical representation of this analysis.

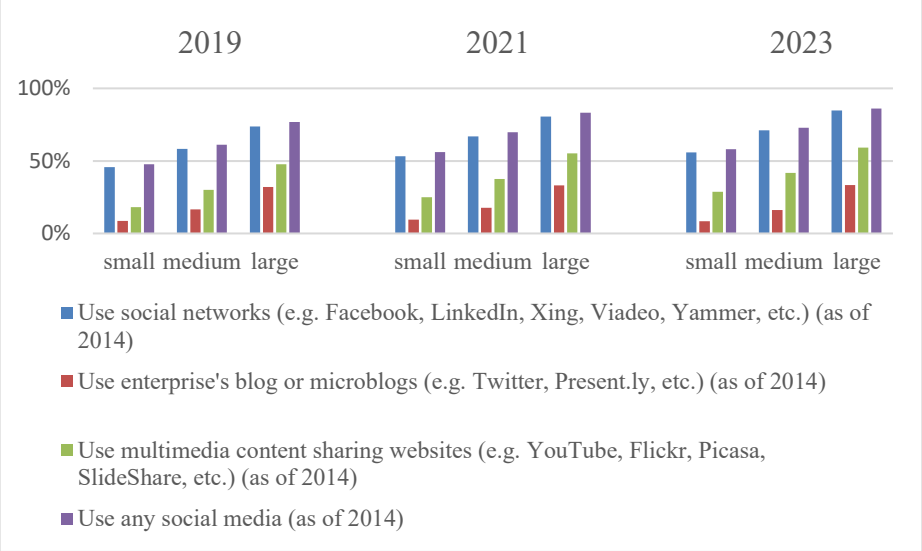


Fig. 3. Social media use by specific purpose and size class of enterprise in the EU countries, Eurostat [9]

4.1 Use of Social Networks

Small Businesses

In 2019, 45.8% of small businesses were using social networks such as Facebook, LinkedIn, Xing, Viadeo, and Yammer. By 2021, this figure had risen to 53.3%, reflecting a growing recognition of the value of social networks for marketing and customer engagement. This upward trend continued into 2023, with 55.9% of small businesses actively using these platforms. This steady increase highlights how small businesses are increasingly leveraging social networks to reach wider audiences and engage with customers more effectively.

Medium Businesses

Medium-sized businesses demonstrated a similar trajectory. In 2019, 58.2% of these businesses were on social networks. This increased to 66.9% by 2021, and further to 71.2% in 2023. The substantial rise over these years underscores the critical role of social networks in the digital strategies of medium-sized enterprises, facilitating broader reach and enhanced engagement with their target markets.

Large Businesses

Large businesses have consistently led the way in social network usage. In 2019, 73.7% of large businesses were using these platforms. This figure grew to 80.6% by 2021 and reached 84.8% in 2023. The high adoption rate among large businesses underscores the

importance of social networks in large-scale business operations and communications, enabling these organizations to maintain a robust online presence and engage with diverse stakeholders.

4.2 Use of Enterprise's Blog or Microblogs

Small Businesses
The usage of enterprise blogs or microblogs among small businesses remained relatively low and stable over the years. In 2019, 8.7% of small businesses utilized these tools. This increased slightly to 9.5% in 2021 but decreased to 8.4% in 2023. Despite the slight fluctuations, the overall low usage indicates that small businesses may find other social media platforms more beneficial for their needs.

Medium Businesses
Medium businesses exhibited a similar pattern. In 2019, 16.6% used blogs or microblogs, increasing to 17.6% in 2021 but dropping slightly to 16.2% in 2023. The stable usage rates suggest that while medium businesses recognize the value of these tools, they may prioritize other forms of social media engagement.

Large Businesses
Large businesses consistently used enterprise blogs or microblogs more than their smaller counterparts. In 2019, 32.0% of large businesses employed these platforms, with a slight increase to 33.2% in 2021 and 33.4% in 2023. The consistent usage indicates that large businesses find value in these tools for detailed communication and content dissemination.

4.3 Use of Multimedia Content Sharing Websites

Small Businesses
The use of multimedia content-sharing websites like YouTube, Flickr, Picasa, and SlideShare has seen a notable increase among small businesses. In 2019, 18.2% of small businesses used these platforms. This rose to 25.0% in 2021 and further to 28.7% in 2023. The increasing trend reflects a growing emphasis on visual and multimedia content as a key component of digital strategies for small businesses.

Medium Businesses
Medium businesses have also increased their usage of multimedia content-sharing websites. In 2019, 30.0% used these platforms, which rose to 37.6% in 2021 and 41.8% in 2023. The significant growth suggests that medium-sized businesses are increasingly investing in visual content to engage audiences and enhance their online presence.

Large Businesses
Large businesses have led in the use of multimedia content-sharing websites. In 2019, 47.6% were utilizing these platforms, with this figure increasing to 55.3% in 2021 and 59.1% in 2023. The high usage rate highlights the importance of visual content in large-scale marketing and communication efforts, enabling large businesses to effectively convey their messages and connect with a broad audience.

4.4 Overall Use of Any Social Media

Small Businesses
Overall social media usage among small businesses has steadily increased. In 2019, 47.6% of small businesses used social media, rising to 56.0% in 2021 and 58.0% in 2023. This consistent growth indicates a broader integration of social media into the operations of small businesses, enhancing their digital engagement and marketing capabilities.

Medium Businesses
Medium businesses have shown a similar trend, with overall social media usage increasing from 61.2% in 2019 to 69.8% in 2021 and 72.8% in 2023. The rising adoption rates reflect a growing reliance on social media platforms to reach and engage with customers, streamline communication, and support business growth.

Large Businesses
Large businesses consistently show the highest overall social media usage. In 2019, 76.9% of large businesses used social media, which increased to 83.3% in 2021 and 86.0% in 2023. The high adoption rate underscores the integral role of social media in their comprehensive digital strategies, allowing large businesses to maintain a strong online presence and effectively engage with their diverse stakeholders.

The Mixed-Effects Model Analysis examines the impact of year and business type on the usage Percentage of various social media platforms while considering the random effect of different categories of social media (e.g., Social Networks, Blogs/Microblogs, Multimedia Sharing). The intercept, i.e. the base level is large businesses in 2019. Among medium businesses, the average usage percentage is 16.17% lower than large companies ($p < 0.001$) and in small businesses, the usage percentage is even lower by 26.87% compared to large companies ($p < 0.001$). The interactions between year and business type were not statistically significant, indicating that the effect of year does not differ significantly across company sizes. The variance between categories (e.g., Social Networks, Blogs/Microblogs) is relatively high, indicating that the usage trends vary significantly depending on the type of social media. There is a clear upward trend in social media usage over time across all company sizes, with significant increases from 2019 to 2023. Larger businesses consistently show higher usage percentages across all categories of social media, with medium and small businesses trailing behind. There is notable variation in how different types of social media are used, but the upward trend is consistent across categories.

These results suggest that while all companies are increasingly adopting social media, large businesses are leading the way, particularly in the use of social networks and multimedia-sharing platforms.

5 Conclusion

Social media's impact on business aligns closely with the rapid development of digital technologies. As digital technologies have advanced, particularly in the realm of social media, they have reshaped how businesses interact with their audiences. Social media's ascendancy as a business catalyst is underscored by its ability to transcend geographical boundaries and facilitate real-time interactions on a global scale. The exponential growth of platforms like Facebook, Instagram, and others has not only provided new channels for marketing and customer engagement but has also revolutionized the way businesses gather and analyze consumer data. This integration of digital technologies into everyday business operations has become pivotal, driving innovation, enhancing efficiency, and opening new avenues for revenue generation. In essence, the evolution of social media mirrors the broader trend of digital transformation, highlighting its crucial role in shaping modern business strategies and fostering economic development globally.

The data analyzed in this paper across various sectors and business sizes in the EU 27 countries illuminates distinct adoption patterns. Small businesses, initially cautious, have progressively embraced social media, with usage rates climbing steadily over the years. Medium-sized enterprises, leveraging their resources, have capitalized on social media's scalability to amplify brand visibility and customer engagement. Meanwhile, large corporations, early adopters of digital strategies, continue to lead in social media utilization, especially in the use of social networks and multimedia-sharing platforms.

References

1. LESDIGIVORES. CH, https://www.lesdigivores.ch/en/the-evolution-of-social-networks-why-your-strategy-needs-to-adapt/
2. F. A. Almazrouei, M. Alshurideh, B. Al Kurdi and S. A. Salloum.: Social Media Impact on Business: A Systematic Review. In: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Hassanien A. E., Slowik A., Snášel V., El-Deeb H., Tolba M. F. (eds.) 697–707 (2020)
3. A. Golmohammadi, D. K. Gauri, and H. Mirahmad. Social Media Communication and Company Value: The Moderating Role of Industry Competitiveness. Journal of Service Research, Volume 26, Issue 1 (2022)
4. Lee D., Hosanagar K., Nair H. S.: Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. Management Science 64(11) (2018)
5. Kietzmann J., Hermkens K., McCarthy P. and Silvestre B.: Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media, Business Horizons Vol. 54(3):241-251 (2011)

6. EUROSTAT, https://ec.europa.eu/eurostat/data/database
7. EUROSTAT Social media use by purpose and size class of enterprise, https://ec.europa.eu/eurostat/databrowser/view/isoc_cismp/default/table?lang=en&category=isoc.isoc_e.isoc_cism
8. EUROSTAT Websites and functionalities by NACE Rev.2 activity, https://ec.europa.eu/eurostat/databrowser/view/isoc_ciwebn2__custom_12011827/default/table?lang=en
9. EUROSTAT, Social media use by type, internet advertising and size class of enterprise, https://ec.europa.eu/eurostat/databrowser/view/isoc_cismpn2__custom_12013754/default/table?lang=en

From Ethics to Liability: Legal Challenges in the Era of Artificial Intelligence

Kristijan Panevski[1], Smilka Janeska Sarkanjac[1], Vladimir Zdraveski[1]

[1]Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Rugjer Boskovic 16, Skopje, 1000, North Macedonia.

Contributing authors: kristijanpanevski@gmail.com; smilka.janeska.sarkanjac@finki.ukim.mk; vladimir.zdraveski@finki.ukim.mk;

Abstract

Artificial Intelligence (AI) is transforming various aspects of society, offering significant benefits but also raising complex legal and ethical challenges. This paper explores these challenges, focusing on AI's legal capacity, responsibility for damages, intellectual property implications, and data privacy concerns. Special attention is given to AI's responsibility as defined by Macedonian law, examining how legal frameworks in Macedonia address the accountability of AI systems. The motivation for this work arises from the urgent need to understand and address the legal and ethical issues posed by AI as it becomes more autonomous and integrated into daily life. By analyzing existing legal frameworks and ethical dilemmas, particularly in the context of Macedonian law, the paper aims to provide insights that can guide the development of policies and regulations. The authors seek to contribute to the ongoing discourse by offering a comprehensive examination of AI's impact on law and ethics, ultimately proposing considerations for a balanced approach that fosters innovation while protecting societal values.

Keywords: Artificial Intelligence Ethics, Legal Regulation of AI, Human Values and AI, AI Liability and Intellectual Property

1 Introduction

At the opening of the Centre for the Future of Intelligence at Cambridge University in 2016, Professor Stephen Hawking famously stated, "The creation of powerful artificial intelligence will be either the best or the worst thing ever to happen to humanity." [1] Discussions about artificial intelligence (AI) began in the 1950s and have become increasingly important over time. The ethical and legal issues related to AI are extensive and span various fields, ranging from personal data protection and liability for damage caused by AI, to questions concerning copyright and intellectual property of AI-generated content and the legal capacity of smart robots, machines, and programs. However, amidst all these complex, often chaotic and contradictory discussions, an inevitable question arises: the role of human values and ethics, and their protection through legal regulation in a future society where AI is an inseparable part of all aspects of human life. The questions raised by AI are urgent, challenging, and provocative because AI is not only a challenge for the law and how it will be regulated, but it also goes much deeper, touching on the ethical question of what it truly means to be human. Alongside the ethical discussions, legal questions are equally important. The rights to personal data protection, intellectual property, compensation for damages, criminal liability, etc., are just some of the areas of law over which AI has, and will increasingly have, a significant impact in the future. These are the issues that will be discussed in this research paper.

The structure of this paper is organized to systematically explore the multifaceted issues surrounding Artificial Intelligence (AI) and its legal implications. In the first chapter, a concise introduction to the topic is provided, highlighting the various questions that arise as AI continues to evolve and integrate into society. This sets the stage for the second chapter, where different definitions of AI are presented, and an attempt is made to synthesize what AI fundamentally represents. The third chapter delves into the legal challenges that AI presents, offering a thorough discussion of the potential legal issues that may emerge. The fourth chapter focuses on AI's responsibility under Macedonian law, examining how AI's accountability is defined and interpreted. This chapter references several key legal acts, providing an analysis of laws related to murder, autonomous vehicle management, and the famous trolley problem, exploring their potential applications and implications in the context of AI. The fifth chapter shifts the focus to intellectual property, exploring how existing laws intersect with AI technologies. In the sixth chapter, the discussion turns to the legal provisions concerning personal data, particularly in how AI systems handle such information. Finally, the seventh chapter offers a comprehensive conclusion, summarizing the insights gained from the previous discussions and reflecting on the broader implications of AI in the legal domain.

2 Definition of AI

The development and use of AI are one of the most exciting topics of our time, and defining artificial intelligence itself is particularly difficult because the term "intelligence" is hard to define. As a result, there are many different definitions. In 2018, the Expert Group on Artificial Intelligence, established by the European Commission,

published a document according to which: "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g., voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g., advanced robots, autonomous cars, drones or Internet of Things applications)." [2] According to one of the reports from the Obama Administration: "Some define AI loosely as a computerized system that exhibits behaviour that is commonly thought of as requiring intelligence. Others define AI as a system capable of rationally solving complex problems or taking appropriate actions to achieve its goals in whatever real-world circumstances it encounters." [3] Microsoft defines AI as "a set of technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways that are similar to what people do." [4] From this, we can conclude that there are two basic AI categories: narrow and general AI. Narrow AI competes with human thinking and reasoning in one area. For example, it can beat the best chess player in the world, but it is not capable of performing other intellectual activities outside of that function. [5] According to Peter Voss, there are two main problems with narrow AI. First, these systems cannot dynamically adapt to new situations, whether it is new situations or new words, expressions, rules, goals, reactions, requirements, etc. [6] In the real world, everything changes constantly, and intelligence is the ability to effectively deal with changes. The second and more serious problem is that AI of this level does not possess its own intelligence that would enable it to think, learn, or solve problems. It actually uses the solution embedded by the programmer for the specific problem the particular system needs to solve. [6] That means that this type of intelligence cannot truly be considered real intelligence. Therefore, these systems are designed to perform certain specific tasks or compete with human thinking and reasoning only in a limited set of situations. This makes these systems capable of performing tasks that require a certain kind of intelligence, but their capabilities are strictly limited to the intended scenarios or specific applications they were programmed for. In essence, narrow AI is specialized for only certain tasks and does not encompass the broad spectrum of intelligent behaviors that are characteristic for humans. Most of the theorists hold the view that there is currently no system that can be called general AI. In fact, it is questionable whether such a system can ever be developed. [5] General AI is a hypothetical concept that has yet not been realized in reality. This is because it requires to create systems with intelligence similar to the human intelligence, which is capable of understanding, learning, and adapting across a whole wide range of domains and situations. These systems could theoretically do any task that requires intelligence, including the tasks that today we believe can only be solved by human thinking, creativity, and emotional intelligence.

3 Legal capacity of AI

As already mentioned, at this point of time AI is at a stage of development in the narrow sense of the word. It means AI is designed to perform specific tasks or to compete with human thinking and reasoning only within a limited scope. This form

of AI is merely a tool for humans to achieve certain goals. However, in the future, we might reach a stage of general AI development. Systems that will easily pass the well-known Turing test [7], evolve, and create and ”live” their artificial ”lives,” even creating feelings similar to human ones. At the moment, AI is not even close to this level of development, so the question of whether or not it should have some of the rights that humans have is hypothetical and philosophical. Firstly, the question is what does it means for a subject of the law to have the ability to hold rights in legal transactions, i.e., to have legal capacity. In the legal theory, legal capacity is considered to be the ability of a natural person to hold legal rights in legal transactions and to have constitutionally and legally guaranteed rights. According to Professor Pop-Georgiev, the main condition for getting legal capacity is the natural person to be born alive and to be born of a woman, meaning to be a human being. [8] This means that a natural person acquires legal capacity at the moment of his or her birth, this capacity is for life, and as such is not conditioned by their age, gender, ethnic, racial, religious, or any other characteristic. It is undisputed that AI does not fulfill any of the conditions above. Namely, it is not a living being from a biological aspect and it is not born by a woman. However, in the law, we already have examples of subjects that are not humans but enjoy the rights granted by law. For an example, animals have certain legal protections to ensure their well-being and humane treatment. Other subjects, which are not humans, such as companies, foundations, associations, and other legal entities, also have the status of holders of rights, such as the right to ownership, the right to judicial protection, fair trial, etc. Given this, it is clear that there is already an established practice in all legal systems that the holder of a right does not have to be a human being. It can be an animal or even legal fictions such as legal entities. Therefore, the question is if AI reaches a sufficiently high level of intelligence and a form of artificial ”consciousness” in the future, could it also be a holder of certain legal rights or enjoy certain legal protections? The answer is yes, AI could have certain rights, but certainly nowhere near the rights that humans enjoy. It is clear that there is a moral obligation to protect AI systems, to design them and stipulate appropriate legal provisions that would protect them from misuse, and to align them with the societal rules that apply. For example, AI would have the right to legal protection in the legal and ethical system. This would be achieved by prescribing rules according to which AI systems would have the right to be designed, and those who design AI would have the obligation to do so considering the positive law and ethical rules, which will lead to the creation of AI in which the general public will have confidence that it is technologically suited for wide use and from which there will be no fears. This would necessarily lead to the creation of rules that will protect AI from improper and unethical use by humans, i.e., protection from human cruelty. Some theorists even go a step further. According to them, if AI systems reach a level similar to human intelligence and consciousness, they would deserve rights from an ethical perspective, such as: [9],[10],[11],[12] - The right not to be turned off against their will; - The right to complete and unobstructed access to their code; - The right to prevent third parties from accessing their code; - The right to copy and reproduce their code; - The right to privacy, primarily the right to refuse to display their ”thoughts.”

4 Responsibility for criminal acts or damages caused by AI use

The question of AI legal capacity is closely connected to the questions of responsibility for criminal acts or damages caused by AI use. Like any human-made product, AI has its shortcomings, whether they are flaws in the design, programming, production process, improper use by users due to insufficient warnings or appropriate usage instructions. These problems can potentially cause harm to third parties, raising the question who will be responsible for that harm. There are no globally accepted opinions regarding the responsibility for criminal acts or damages caused by AI, so the answer should be looked for in the laws of the country where the analysis is to be done. For the purposes of this paper, the law of the Republic of North Macedonia will be considered as the applicable law. There are two main answers regarding the question who should bear the responsibility for a criminal offense or damage caused by AI. The first is that the responsibility should be born by the AI system. However, as previously mentioned, the current level of AI development is not high enough to be accountable for offenses committed by AI. On the other hand, considering the level of AI development, it is more logical to consider it just as a tool, a certain product, used by or acting for a certain person, and who will bear the responsibility in that case will be determined in the following pages. If we accept that AI represents a product that does not have its own personal responsibility and legal capacity, in the case of a criminal offense or damage caused by using AI, the provisions in the Criminal code and the Law on obligations can be applied to determine the liability for the damage.

4.1 Responsibility for criminal acts

In the simplest terms, criminal responsibility, or guilt, is the responsibility for committing a criminal offense as defined in the Criminal code of the Republic of Macedonia (CC). [13] Criminal responsibility is determined in a criminal proceeding to impose an appropriate penalty for the committed offense. [14] Regarding the criminal responsibility for committing a criminal offense, there are numerous provisions in the CC whose application depends on the type and manner in which the criminal offense was committed. None of these provisions explicitly mention AI, but in any case, if the offense is committed using AI (or any other tool or product), the user, and in certain cases, the manufacturer of this technology will be held accountable for such an offense. As an example, in this section, we would consider the criminal offense - "Murder". Namely, if death occurs to a person due to the use of AI, the general provision of Article 123, paragraph 1 of the CC would apply, which states: "(1) Whosoever deprives another of life shall be sentenced to at least five years of imprisonment." Depending on the reasons and the manner in which this offense was committed, shorter or longer prison sentences are also foreseen. Thus, for example, a prison sentence of at least 10 years to life imprisonment is foreseen if the act was committed in a cruel or insidious manner, if the offense deliberately endangered the life of another person, if it was done for gain, to carry out or conceal another criminal offense, out of reckless revenge or other low motives, to take the life of a woman known to be pregnant or a minor, etc. On the other hand, if the murder is committed out of negligence, a shorter prison

sentence is foreseen, ranging from six months to five years. It is noticeable that in the Macedonian legal system, there is a clear distinction in the severity of the offense and the length of the penalty depending on the motives for which the offense was committed and the manner in which it was committed, on which the degree of guilt depends. In the Macedonian legal system, but also in other legal systems, there are two main degrees of guilt in the criminal law: "Intent" and "Negligence". According to Article 13 of the CC: "Premeditated crime shall be considered when the offender is aware of his offense and wanted its commission or when the offender was aware that due to its commission or non-commission harmful consequence may appear as a result and yet he approved its occurrence." This means that the perpetrator must have had a conscious and intentional decision to commit the criminal offense. He was aware of his actions and the consequences they could cause and wanted to commit that offense or accepted the possible harmful consequences of his action or inaction. The qualification of criminal offenses committed with this degree of guilt is more severe, so accordingly, the penalty for the perpetrator is higher. On the other hand, Article 14 of the CC states: "A crime committed due to negligence shall be considered when the offender was aware that the commission or non-commission of the crime may result in harmful consequence, but has lightheartedly considered that he might prevent it or it may not occur or was not aware of the possibility for resulting in harmful consequence, although due to the circumstances and according to the personal characteristics he might have been aware of that possibility." In negligence as a degree of guilt, the perpetrator's intention to commit the criminal offense is absent. The perpetrator is still responsible because he was aware that the offense could occur, although he recklessly believed that it would not. Given the absence of intent, shorter penalties are foreseen for criminal offenses committed under this degree of guilt. Taking into account the degrees of guilt and the criminal offenses given earlier as examples, we can conclude that in a situation where the perpetrator of the offense is the user of the tool with which the offense was committed, such as a firearm, vehicle, or some other object or, in this case, a specific type of AI system, and he committed that offense directly, by his guilt, with intent or out of negligence, he himself is responsible for the committed offense and would be expected to be held accountable for that offense according to the provisions of the CC. So, if someone performs a certain action using AI, which directly causes death to a third person due to his guilt, unless conditions for imposing a suspended sentence are met, he will receive a prison sentence of 5 years at best, and at worst, life imprisonment. The problem arises when the damage is caused by negligence, and it is not directly the user's fault, but results from the way the AI system is programmed or produced, or the user was given incorrect instructions for use. In that case, the responsibility cannot be sought from the user of the system, but should be located with the legal entity that put the AI system into use or produced it, which caused the specific offense. The responsibility in this case would be with the seller or manufacturer of the AI system, depending on the type of product, and their degree of guilt would typically be negligence unless proven otherwise. The best example of such responsibility of the manufacturer or seller of the AI system would be a traffic accident with fatal consequences, and the responsible party for the occurrence of the traffic accident is an autonomous vehicle, i.e., a vehicle that moves independently through

the use of an AI system. Namely, these are vehicles equipped with AI technology that allows them to move without human assistance. The goal of developing autonomous vehicles is to improve traffic safety (analyses show that 90-95 percent of traffic accidents are the result of human errors), as well as to increase the efficiency of vehicle management and ease travel, especially for people who cannot or do not want to drive themselves. [15] However, no matter how advanced such technology is (currently, there are no fully autonomous vehicles), it is certain that in practice, errors will occur in the future that can and will lead to serious consequences for people's lives and health. If death or bodily injury to a person occurs as a result of an error in the driving system of an autonomous vehicle, the question is who would be criminally responsible for it. In the Republic of North Macedonia, it is currently not allowed to operate a vehicle unless the driver has full control over the vehicle, so if a criminal offense occurs using a vehicle with such a system, the person who was supposed to control the vehicle would undoubtedly be held accountable. However, with the increasing development of this technology and the growing number of such vehicles in the Republic of North Macedonia, there is a possibility that the use of fully autonomous vehicles will be allowed in the future. In such conditions, it is clear that the entire responsibility cannot be placed on the owner of the vehicle or the person who could have operated it. Especially since the manufacturers of such vehicles are increasingly convinced of their ability to drive independently and safely on the roads without human assistance and have publicly expressed that. [16] One of the options is for the "driver" to agree to "shared" responsibility with the manufacturer of the vehicle or even take full responsibility in case of a traffic accident before the vehicle starts moving. When deciding whether to take responsibility, the driver will primarily have to consider that when programming such systems, in addition to unforeseen errors that may occur, manufacturers will also have to consciously include certain rules that the vehicle will need to follow in specific situations where damage may occur. These are ethical decisions that should guide the vehicle on how to act in a given situation. Namely, should the vehicle prioritize the safety of passengers or consider all circumstances to minimize harmful consequences, even at the cost of endangering the safety and lives of the passengers inside it? The question arises, who would buy or use a vehicle if they knew there was a possibility of being "sacrificed" to protect someone else's safety outside the vehicle. One of the most well-known ethical questions is the trolley problem. The trolley problem is a thought experiment in ethics. In principle, the problem is as follows: "You are driving a trolley that is out of control and heading towards five people working on the track you are on. To the right is another track where you can redirect the trolley. If you pull the lever, the trolley will switch tracks, and the five people on the track you were previously on will be saved. However, on the side track is one person who also cannot avoid the oncoming trolley. You have two options: 1. Do nothing and let the trolley kill five people on the main track. 2. Pull the lever and redirect the trolley to the side track, killing one person. Which option is ethically correct?" [17] Currently, the software used in autonomous vehicles does not have a solution to these problems. The system is programmed to react to any obstacle, without considering what the obstacle is or their quantity. An exception to this rule is the size of the obstacle, so if it has to choose between two obstacles, the vehicle would be directed towards the smaller one.

[15] In such a situation, in the case of full autonomy of the vehicle and the possibility of errors in its operation, as well as the theoretical possibility of systematically embedding certain rules of action, as in cases like the trolley problem, which are unclear whether they are correct or not, there would be a greater degree of skepticism and resistance for the driver to take responsibility for damage if the control of the vehicle is entirely with the AI system. In our opinion, in such a case, the responsibility would lie with the manufacturer of the AI system, i.e., the autonomous vehicle. Manufacturing and putting autonomous vehicles into use undoubtedly require enormous funds, so it is clear that only legal entities, i.e., corporations involved in this type of business activity, could do this. In the Macedonian legal system, in addition to natural persons, criminal liability is also foreseen for legal entities. According to Article 96-a of the CC: "(1) A fine shall be imposed as main sentence for crimes of legal entities. (2) The fine shall be imposed in an amount that cannot be less than 100.000 Denars nor more than 30 million Denars." Depending on the severity of the offense, the manner of its commission, and its consequences, there is a possibility of imposing lighter or harsher penalties. [13] Considering this, if a person is killed as a result of an autonomous vehicle, and the vehicle is to blame, then criminal proceedings will be initiated against the legal entity, and an appropriate penalty will be determined. These provisions can be applied not only to situations like the example with autonomous vehicles, but also to any criminal offense caused by an error in the functioning of the AI system that is not the fault of the user or a third party or due to certain deliberately embedded rules that result in harmful consequences, for which criminal liability is foreseen.

4.2 Civil liability for compensation of damages

Liability for compensation of damages refers to the obligation of the person causing the damage (the wrongdoer) to compensate the injured party. The general provisions for compensation of damages are contained in the Law on obligations of the Republic of North Macedonia. [18] According to Article 141 of the Law on obligations: "Anyone who causes damage to another through their fault is obliged to compensate for it. For damages caused by objects or activities that pose an increased risk of harm to the environment, liability is established regardless of fault."

From these provisions, it can be determined that there are two types of liability for compensation of damages: - Subjective liability: Liability due to fault (intentional or due to negligence). - Objective liability: Liability regardless of fault (as the owner of a dangerous object or performer of a dangerous activity). In addition to the existence of liability, legal theory stipulates that for an obligation to compensate damages to exist, the following three conditions must be met: [19] - Existence of an unlawful harmful action: An action that causes damage to the injured party. The damage can be caused by an action, omission, or through an object or activity that represents a source of increased danger. The unlawfulness of the action means that it violates an imperative legal norm. - Occurrence of damage: The damage is a consequence of the harmful action. According to Article 142 of the Law on obligations: "Damage is the reduction of someone's property (ordinary damage) and the prevention of its increase (lost benefit), as well as the violation of personal rights (non-material damage)." [18] - Causal link: The damage must be a consequence of the unlawful harmful action, i.e., the damage

must result from the action or omission of the wrongdoer or originate from a dangerous object or activity. Similar to the previous point, the liability for compensation of damages caused by the user of an AI system (the wrongdoer) will lie with the user if they intentionally or with insufficient attention, i.e., due to negligence, caused damage to another person (the injured party) while using the AI system. However, the question arises as to who will compensate for the damage if the user of the AI system is not at fault for the harmful action that caused the damage. To better illustrate this, let's use the same example of autonomous vehicles. Who will be responsible for compensating the damage caused by an error in the AI system for autonomous driving or due to actions taken by the AI system based on embedded rules from the manufacturer? At first glance, the answer seems simple. According to the Law on compulsory insurance in traffic (LCIT), [20] "Vehicle owners are obliged to conclude a compulsory insurance contract with an insurance company before allowing the vehicle to operate in traffic." Regarding damage caused by a motor vehicle, "the injured party has the right to file a claim for damage compensation directly with the responsible insurance company." Given the above, it is clear that if the use of fully autonomous vehicles is allowed on our roads, they must be insured. In case of damage caused by an autonomous vehicle, the damage should be compensated by the insurance company. It is up to the insurance companies, in cooperation with the National Insurance Bureau of Macedonia (NIB) and the Insurance Supervision Agency, along with relevant state authorities, to develop strategies and establish appropriate rules on how and in what manner these vehicles will be insured. Considering that these are a specific category of vehicles, there is a possibility that the aforementioned entities might decide to set a higher insurance premium for these vehicles to cover the increased risks of damage. Or, they might determine that these vehicles (despite all risks of system errors) are safer than human-operated vehicles and might reduce the insurance premium for such vehicles to encourage their greater use. However, what if the autonomous vehicle is not insured and is not operated by a driver, yet causes damage due to a system error or the before mentioned embedded rules for AI system reactions in certain situations? The injured party can directly demand compensation from the owner of the autonomous vehicle. Despite the lack of fault on the owner's part, they may not even be in the vehicle, and the vehicle might be unoccupied, the owner is typically liable for damage caused by their vehicle, as it is inherently a dangerous object. This is supported by to Article 159 of the Law on obligations, according to which: "For damage related to an object, movable or immovable, whose position, use, property, or mere existence poses an increased risk of harm to the environment (dangerous object) or an activity that poses an increased risk of harm to the environment (dangerous activity), it is presumed that the damage originates from that object or activity, unless it is proven that the cause lies with the injured party or a third party, or due to force majeure."

Also, according to Article 160 of the Law on obligations: "For damage caused by a dangerous object, its owner is liable, and for damage caused by a dangerous activity, the person engaged in it is liable."

However, there is an exception to this rule, provided in Article 163, paragraph 2 of the Law on Obligations, which states: "The owner of the object is exempt from liability if they can prove that the damage was caused solely by the action of the injured party

or a third party, which they could not foresee and whose consequences they could not avoid or eliminate." Therefore, in a potential court proceeding for compensation initiated by the injured party, the owner of the autonomous vehicle will have the right to prove that, despite the fact that the damage was caused by their autonomous vehicle, the fault for the damage lies with a third party, namely the manufacturer of the autonomous vehicle. If they prove that the damage was caused due to a system error or pre-installed rules, i.e., due to the manufacturer's fault, the owner will be exempt from liability and will not be obliged to compensate the injured party. In such a case, the injured party will need to direct their claim towards the manufacturer of the autonomous vehicle.

5 AI and intellectual property

The main issues related to AI and intellectual property arise from the way generative AI technology impacts the creation of new works. These include questions regarding the protection of intellectual property for AI models, the rights of authors of works created with generative AI systems, as well as the protection of the rights of authors of existing works used to train AI models.

5.1 Protection of intellectual property for AI models

In the Macedonian law, the protection of intellectual property is mainly realized through patents and copyrights. According to Article 25 of the Law on industrial property: [21] "(1) A patent shall protect an invention in all technology fields, if new, if has an inventive contribution, and if applicable in the industry. (3) The following shall not be considered to be an invention in terms of paragraphs (1) and (2) of this Article: 3) a plan, rule and procedure for performing an intellectual activity, for games or for performing business activities, as well as a computer program;" Having this in mind, it is clear that an AI model, written as a computer program cannot be considered an invention and as such cannot be protected with a patent. On the other hand, according to Article 12 of the Law on copyright and related rights: [22] (1) A copyrightable work, within the meaning of this law, is an intellectual and individual creation in the field of literature, science, and art, expressed in any manner and form. (2) Particularly considered as a work of authorship are: (3) computer program, as a written work;

Also, according to Article 95 of the Law on copyright and related rights: "A computer program, within the meaning of this law, is a program in any electronic form of expression, including preparatory material for its creation, provided it is an individual and intellectual creation of its author." Considering the aforementioned, it is clear that in the Macedonian legal system, an AI model can be protected as a copyright, and the author of that copyright will have all of the rights stipulated in the Law on copyright and related rights.

5.2 Protection of content generated with AI

When it comes to the protection of AI-generated content under the Macedonian Law on copyright and related rights, the general rule is that it can be protected, but only

if it meets the criteria for being considered an "individual and intellectual creation." Having that in mind, for a work generated with AI to be protected, the petitioner must prove that it is an original work that involved the individual and intellectual effort of the human author. If the work is generated fully autonomously the protection will be impossible, since the law recognizes only human authorship. However, if the petitioner proves that the AI generated content is directed or significantly influenced by a human author, it is possible to attribute the authorship to the human author, thus making the work eligible for protection under the Macedonian Law on copyright and related rights.

5.3 Protection of the rights of authors of existing works used to train AI models

The popular deep learning models are trained with large amounts of data collected from the internet, some of them copyrighted works. This can lead to the violation of the author's right to control the reproduction of their work unless the processing of these protected data falls under the exceptions provided by the relevant legislation. Additionally, the use of works generated by AI can also lead to violations of the copyright of the author who created the work on which the model was trained. [23] According to the Macedonian Law of copyrights and related works, consent from the authors is generally required for their use in training AI models. According to the Macedonian Law on copyrights and related works, the use of copyrighted works is governed by the rights and permissions of the authors. Namely, according to Article 26 paragraph 2 of the Law of copyrights and related works: "(2) The author has the exclusive right to permit or prohibit the use of their work, or its copies, by others, except in cases specified by this law."

Also, Article 27 of the Law of copyrights and related works details the specific material rights of the authors, including: - Reproduction: The right to make copies of the work. - Distribution: The right to distribute copies of the work to the public. - Public communication: The right to communicate the work to the public, such as through broadcasting or public performance. - Adaptation: The right to modify or adapt the work.

These rights are owned by the author, and he decides whether to grant permission for their work to be used for AI model training. Using their work without permission is a violation of the law and grounds for the author to seek compensation for the unconsented usage of their work. The law allows certain uses of copyrighted works without the author's consent only under specific conditions, such as for private use, educational purposes, or public safety (Articles 51 and 52). However, these exceptions do not typically cover the use of works for training AI models, which would generally be considered a commercial or industrial use.

6 Misuse of personal data by AI systems

The misuse of personal data by AI systems does, and in the future will presents a lot of significant dangers, especially those concerning the privacy and security of the data subject. According to the Law on personal data protection [24] of the Republic

of North Macedonia, personal data always must be handled with the highest degree of care, in order to ensure compliance with legal standards and to protect individuals' rights. Namely, in order to protect the personal data and prevent misuse, Article 9 of the Law on personal data protection stipulates the following principles: "(1) Personal data shall be: - processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency'); - collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;"

However, when these principles are not respected, the most common and likely dangers are: - Unauthorized access and data breaches: If the AI systems are not properly secured, they can become targets for cyberattacks and can lead to unauthorized access and data breaches. Article 36 of the Law on personal data protection emphasizes the need to take adequate security measures which should lead to the protection of the personal data during processing. All of this in order to prevent unauthorized access and data breaches. - Automated decision-making and profiling: One of the biggest dangers when using AI systems involve the automated decision-making process and profiling, which can have significant impacts on individuals. Article 26 of the Law grants the individuals with the right not to be subjected to decisions based solely on automated processing, including profiling, unless specific conditions are met, such as explicit consent or legal necessity. - Violation of data subject rights: The AI systems must comply with the most important principles of data processing, which include lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality. Article 9 determines these principles, emphasizing that personal data must be processed in a manner that ensures appropriate security and protection against unauthorized or unlawful processing. - Lack of consent: The processing of personal data without the proper consent of the data subject is a major concern. Article 11 requires that consent be obtained for processing personal data, and individuals must be informed about the processing purposes and their rights. - Impact assessments and prior consultation: Before deploying AI systems that process personal data, organizations must conduct data protection impact assessments to identify and mitigate risks. Article 39 stipulates the impact assessments for processing activities that are likely to result in high risks to individuals' rights and freedoms. Additionally, Article 40 requires prior consultation with the Data Protection Agency when high-risk processing is identified.

7 Conclusion

Artificial Intelligence (AI) represents a powerful tool for technological innovation, but it also introduces complex legal, ethical, and societal challenges. To effectively address these issues, proactive measures must be taken to mitigate the potential negative impacts of AI. First and foremost, clear and robust legal frameworks need to be established to define AI's legal capacity and responsibility, particularly in areas such as liability for damages, criminal accountability, and intellectual property rights. In the context of Macedonian law, for example, there should be specific legal provisions

that outline the responsibility of AI systems, ensuring that they are held accountable in a manner consistent with human operators.

To address the misuse of personal data, stronger data protection regulations must be implemented, with stringent enforcement mechanisms to safeguard individual privacy rights. Additionally, there should be ongoing efforts to build human and infrastructural capacities to manage and regulate AI effectively, including the training of legal professionals and technologists to navigate the evolving landscape of AI-related issues.

Moreover, ethical guidelines should be established to govern the development and deployment of AI, ensuring that it aligns with societal values and respects human rights. These guidelines should be informed by continuous monitoring and research to adapt to new challenges as AI technologies advance.

Finally, international collaboration is essential to harmonize legal and ethical standards across borders, addressing the global nature of AI's impact. By taking these steps, we can ensure that AI serves as a force for good, advancing human rights and societal well-being while minimizing risks. As we move forward into this new era, it is imperative that we approach AI with responsibility and foresight, using it to solve pressing global challenges and enhance the quality of life for all.

References

[1] Hawking, S.: The best or worst thing to happen to humanity. Speech at Leverhulme Centre for the Future of Intelligence, Cambridge (2016)

[2] AI, H.: High-level expert group on artificial intelligence. Ethics guidelines for trustworthy AI **6** (2019)

[3] House, W.: Preparing for the future of artificial intelligence. Executive Office of the President National Science and Technology Council. Committee on Technology (2016)

[4] Smith, B., Shum, H.: The future computed. Artificial Intelligence and its role in society (2018)

[5] Dempsey, J.X.: Artificial intelligence: An introduction to the legal, policy and ethical issues. USA, Berkeley: Berkeley Center for Law & Technology (2020)

[6] Voss, P.: From narrow to general ai. Institution Machine, Retrieve from https://medium. com/intuitionmachine/from-narrow-to-general-ai-e21b568155b9 (2017)

[7] Turing, A.M.: Computing machinery and intelligence. Springer (2009)

[8] Pop-Georgiev, D.: Civil Law: General Part, Property Law. University of Ss. Cyril and Methodius in Skopje (1966)

[9] Tavani, H.T.: Can social robots qualify for moral consideration? reframing the

question about robot rights. Information **9**(4), 73 (2018)

[10] Schwitzgebel, E., Garza, M.: A defense of the rights of artificial intelligences. Midwest Studies in Philosophy **39**(1), 98–119 (2015)

[11] Darling, K.: Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. Edward Elgar Publishing (2016)

[12] Levy, D.: The ethical treatment of artificially conscious robots. International Journal of Social Robotics **1**(3), 209–216 (2009)

[13] Criminal code. Official Gazette of the Republic of Macedonia No. 37/96

[14] Kambovski, V.: Criminal Law – General Part. University of Ss. Cyril and Methodius in Skopje (2011)

[15] Bartneck, C., Lütge, C., Wagner, A., Welsh, S.: An introduction to ethics in robotics and AI. Springer (2021)

[16] Kirsch, D.A., Chowdhury, M.A.: Fanbois and fanbots: Tesla's entrepreneurial narratives and corporate computational propaganda on social media. World Electric Vehicle Journal **14**(2), 43 (2023)

[17] Thomson, J.J.: The trolley problem. Yale LJ **94**, 1395 (1984)

[18] Law on obligations. Official Gazette of the Republic of Macedonia No. 18/2001

[19] Galev, G., Dabović Anastasovska, J.: Obligations Law. University of Ss. Cyril and Methodius in Skopje (2021)

[20] Law on compulsory insurance in traffic. Official Gazette of the Republic of Macedonia No.88/2005

[21] Law on industrial property. Official Gazette of the Republic of Macedonia No.21/2009

[22] Law on copyright and related rights. Official Gazette of the Republic of Macedonia No.115/2010

[23] Cuntz, A., Fink, C., Stamm, H.: Artificial intelligence and intellectual property: An economic perspective. World Intellectual Property Organization (WIPO) Economic Research Working Paper Series (77) (2024)

[24] Law on personal data protection. Official Gazette of the Republic of Macedonia No.42/2020

The impact of custom Salesforce applications on resource visibility utilization and data security

Melina Stojanoska[1], Boro Jakimovski[1],
Smilka Janevska Sarkanjac[1], Eftim Zdravevski[1], Petre Lameski[1]

[1]Ss Cyril and Methodius University in Skopje, Faculty of Computer Science and Engieering,Ruger Boskovik 16, Skopje, 1000, Macedonia.

Abstract

This study evaluates a custom Salesforce application designed to enhance organization resource management. The application monitors resource availability, project assignments, and skill sets by deploying Apex code, streamlining these processes with intuitive color coding and a user-friendly interface. Additionally, it examines the cybersecurity implications of using Salesforce for resource management. The research aims to demonstrate how such custom applications can improve visibility, allocation, and utilization of resources, thereby boosting overall project management efficiency. The hypothesis suggests that this Salesforce integration will substantially elevate resource management by centralizing data and providing real-time insights, ultimately facilitating more informed decision-making.

Keywords: CRM, Data Security, Resource Optimization

1 Introduction

Customer relationship management (CRM) systems started their envisioning, design, and development in the 1970s due to the paradigm shift of businesses from product-oriented to customer-oriented. Peter Ducker states, "The purpose of business is to create and keep a customer.". This saying emphasizes the importance of managing customer relationships; as such, CRM systems have reached a pivotal role, giving a significant advantage to companies that adopt them.

CRM software offers many benefits. It automates sales and marketing processes, reduces manual work, and helps businesses manage tasks such as cold calling, customer

management, and follow-up, linking sales and marketing activities for better efficiency. CRM can improve overall business processes when integrated with an ERP (Enterprise Resource Planning) system, creating greater synergy and efficiency. In addition, CRM provides valuable business intelligence and data mining from customer databases. Despite some challenges, implementing a CRM solution generally leads to increased productivity, better efficiency and improved integration of different processes.

Salesforce is a cloud-based CRM software that revolutionizes how companies interact with customers, streamline processes, and enhance service delivery at scale. By aggregating data from various sources, Salesforce provides a unified view of customer information across sales, service, marketing, commerce, and IT departments through its comprehensive suite of products known as Einstein 1. Powered by artificial intelligence, this integration significantly boosts productivity and enables personalized customer experiences that companies strive for [1].

Since its inception in 1999 by former Oracle executive Marc Benioff and his cofounders, Salesforce has championed the software-as-a-service (SaaS) model, allowing users to access its services via a web browser without needing cumbersome client software installations. This model has contributed to Salesforce's exponential growth, evidenced by its impressive $31.3 billion revenue in the fiscal year ending January 31, 2023, marking an 18% increase from the previous year[2]. Strong integration capabilities, customer-centric support services, robust security, and AI-powered insights are some other attributes that make Salesforce a better CRM than other available options in the market [3].

Security within Salesforce is becoming increasingly challenging as organizations recognize the need to manage security within SaaS applications. Traditionally, companies have focused on auditing potential partners but often overlooked the platforms themselves. The Salesforce ecosystem typically includes a mix of internally developed and third-party applications, making it essential for organizations to understand all apps interacting with their data and implement appropriate security measures.

As Salesforce becomes deeply integrated into an organization, the number of users with access can grow, making it difficult to control and track access levels. This issue is exacerbated as employees join, move within, or leave the company. Without stringent access controls, internal employees can pose significant security risks. Simple actions like downloading data to a spreadsheet can create immediate and long-term security issues. Organizations must ensure that Salesforce access is restricted to necessary personnel and that permissions are managed to prevent unintentional harm.

There is a rising trend of citizen developers within organizations. Forrester research indicates that 39% of surveyed digital and IT professionals allow employees outside of IT to develop applications. While this can be beneficial for quickly meeting business needs, it also introduces security risks. Often, low-code developers are treated differently from full-stack developers, which can lead to security vulnerabilities. Even in low-code environments, citizen developers can make mistakes that compromise security. Organizations need to establish safeguards to ensure these developers can operate securely. Salesforce's broad applicability can lead to internal conflicts about whether Salesforce security should be managed by IT, sales operations, or another department. While the CISO is responsible for overall security, the actual implementation of

security measures often falls to developers or program managers. Organizations must clarify ownership and responsibilities to avoid confusion and ensure adequate security management.

Machine identities, including non-human identities such as automated processes or applications, are common in SaaS and other applications and must be handled cautiously. These identities complicate authentication and data flow control between applications, creating potential vulnerabilities. Organizations need a comprehensive understanding of all machine identities in use and the security protocols associated with them [4].

Salesforce often houses an organization's most sensitive data. Although the CRM tool significantly helps maintain security, it's crucial to remember that combating cyber threats is a shared responsibility. Both in-house security teams and Salesforce administrators must protect company data, including managing access and continually monitoring data activity. As threats evolve, it's important to understand current security risks. These risks include complex permission models leading to data exposure, sensitive data being stored in inappropriate locations, APIs inadvertent leaking information, public misconfigurations making data accessible, and a lack of coordination between Salesforce admins and security teams like CISOs or CIOs. Salesforce professionals and security teams can employ several strategies to mitigate these risks. Starting with a comprehensive data risk assessment helps identify potential vulnerabilities.

The rise of digital-first strategies has increased the use of Salesforce communities and experience sites. These sites, though useful, pose high data exposure risks. Misconfigured permissions can lead to sensitive data being publicly accessible. Regular security reviews of permissions can prevent such exposures. Salesforce also allows users to share files and attachments through public URLs. Many users are unaware of this setting and its security implications. Organizations should review and control who can create public links to prevent unintended data sharing [5]. Salesforce boasts several key security features that ensure robust protection for its users. Firstly, the physical security of its data centers is safeguarded with industry-standard measures, including biometric access controls, video surveillance, and perimeter fencing. Network security is fortified with multiple layers, such as firewalls, intrusion detection and prevention systems, and continuous network traffic monitoring.

In terms of application security, Salesforce incorporates built-in controls to protect against common web application attacks like cross-site scripting (XSS), SQL injection, and phishing. Developers can also utilize tools provided by Salesforce to build secure applications following industry-standard security practices. Data security is another critical aspect, with multiple layers of protection, including data encryption in transit and at rest, access controls based on roles and permissions, and continuous monitoring and auditing of data access. User access control is managed through various tools such as multi-factor authentication, role-based access control, and permission sets. Salesforce's identity and access management (IAM) system integrates with popular identity providers like Active Directory and Okta.

Additionally, IP whitelisting restricts access to Salesforce instances to specific IP addresses, preventing unauthorized access from outside the organization's network.

Role-based access control (RBAC) allows organizations to define access permissions based on individual user roles, ensuring that users only access the data necessary for their job functions. Data at rest and in transit is protected through encryption, preventing unauthorized access to sensitive information. Data masking is another security feature that involves hiding sensitive data by replacing it with fictitious data, thus protecting it from unauthorized access by users who do not need to see it. Salesforce also provides extensive monitoring and logging capabilities, enabling organizations to track user activity and identify security incidents promptly. Regular security assessments help organizations identify and address vulnerabilities before they can be exploited.

Salesforce facilitates a comprehensive security model designed to meet the needs of even the most security-conscious organizations. However, it is essential to note that security is a shared responsibility between Salesforce and its customers, who must ensure their use of Salesforce complies with industry-standard security practices. Salesforce offers a flexible and granular record-level security model that allows organizations to restrict access to individual records based on user roles, permissions, and criteria. This model includes three main layers: Organization-Wide Default (OWD) settings, Role-Based Access Control (RBAC), and Record-Level Security. OWD settings define the default level of access to records for users without specific record-level access permissions and can be configured at the organization, object, or individual record level. RBAC manages access to records based on user roles and hierarchies, using a combination of user roles, profiles, and sharing rules. Record-level security controls access to individual records based on criteria such as record ownership, field values, or record types, using sharing rules, manual sharing, and criteria-based security [6].

Safeguarding data privacy and security is critical for maintaining customer trust and achieving success. Salesforce, a leading CRM platform, provides essential tools and policies to help businesses implement effective controls for data privacy. Salesforce offers numerous certifications for its applications, underscoring its commitment to data security. The company's trust program ensures compliance with various data privacy standards, guaranteeing that customer data is secure and trans-missions are protected. Enterprise users can access a broad range of data security controls based on their Salesforce product. These controls include Three-Dimensional Secure (3D Secure), Data Filtering, Geographic Containment, IP Geo-Fencing, Access Management, Encrypted Modeling, and Intrusion Prevention Systems (IPS). Such measures enable enterprises to protect their data comprehensively. To ensure customer data privacy, Salesforce employs industry-standard protocols such as Secure Socket Layer/Transport Layer Security (SSL/TLS), Data Leakage Prevention (DLP), Strong Authentication, Access Management Solutions, Encryption, and Data Segmentation. These protocols provide robust security assurances for customer data. Salesforce also offers advanced enterprise search and data analysis technologies, enabling businesses to gain insights while ensuring data security and regulatory compliance. The CRM and Salesforce Platform Security features enhance access control and prevent fraud. Additionally, Salesforce provides solutions for release management, data backup and recovery, and comprehensive security measures for its enterprise customers. Release management prevents data over-disclosure during new technology implementations. Data backup and recovery solutions ensure that customer data is recoverable in case of

loss, deletion, or corruption. The security solutions encompass authentication, access control, encryption, surveillance, data protection, vulnerability scanning, and threat detection and prevention.

In summary, Salesforce delivers the necessary tools and controls for enterprise customers to maintain customer data privacy and security. The extensive certifications, protocols, analytical technologies, and security solutions available reinforce this commitment [7]. Enhancing Salesforce security begins with establishing a robust system of roles and profiles, ensuring users have only the necessary access and thus reducing the risk of internal data breaches. Flexible access control can be achieved through permission sets, which allow for additional permissions without altering core profiles, making them ideal for complex setups. Implementing Multi-Factor Authentication (MFA) adds an extra layer of security by requiring users to verify their identity in multiple ways, significantly reducing the likelihood of unauthorized access, particularly from external sources. Effective data protection involves applying Data Loss Prevention (DLP) strategies such as encryption and tokenization to safeguard sensitive data in Salesforce against unauthorized access and leaks. Regular monitoring and auditing of user activity are crucial for identifying potential security risks and ensuring compliance with security policies. Additionally, Salesforce's Security Health Check tool helps evaluate the platform's security settings, proactively identifying and addressing vulnerabilities.

Integrating third-party applications with Salesforce requires careful consideration of the app provider's security standards and alignment with the organization's security policies. Implementing the principle of least privilege by granting only the necessary permissions for these apps minimizes the risk of unauthorized access or data leaks. Monitoring third-party app integrations ensures ongoing compliance and security standards are maintained. Cultivating a security-first mindset within the organization is essential. Emphasizing the importance of protecting sensitive data and adhering to security policies, alongside hands-on training sessions for employees to navigate Salesforce's security features effectively, helps prevent breaches caused by user errors. Staying informed about the latest updates and vulnerabilities is crucial. Subscribing to Salesforce security alerts and engaging in community forums helps keep the environment secure by learning from others and staying current with security practices. By diligently applying these strategies, the Salesforce platform's security is significantly strengthened, protecting the organization's data and maintaining customer trust [8].

This research investigates how Salesforce can enhance resource management by developing a custom application. The focus is on improving the tracking, allocation, and utilization of organization resources. Utilizing Apex code, the custom Salesforce application will monitor resource availability, project assignments, and skill sets. It will present data through intuitive color coding and a user-friendly search component to optimize resource management processes. Additionally, this research covers the cybersecurity aspects of leveraging Salesforce for resource management.

By evaluating the impact of this application, the study aims to provide insights into how Salesforce can effectively enhance resource allocation, improve visibility into resource utilization, and ultimately contribute to more efficient project management. This research seeks to contribute valuable findings to the field, highlighting the

potential of CRM platforms like Salesforce in supporting organizational efficiency and strategic resource optimization.

The research questions that we try to assess in the study are as follows:

- RQ1: Is the impact on resource utilization, availability and skill set visibility from using custom Salesforce applications positive or negative?
- RQ2: Can UI elements used in custom applications capture resource allocation complexities?
- RQ3: Does Salesforce as CRM system cope with modern cybersecurity challenges?

This study hypothesizes that implementing a custom Salesforce application will significantly improve resource management efficiency by centralizing data, providing real-time insights, and facilitating informed decision-making in a secure and reliable manner. The paper is organized as follows: In section 2 we give overview of similar studies that can be found in the literature. In section 3, we describe the research methodology. After that, in section 4, we give an overview of the results, and finally, in section 5, we give a short conclusion.

2 Related Work

Previous studies have underscored the benefits of integrating CRM systems with project management tools, aligning with the proposed research on leveraging Salesforce for resource management. The research conducted by the authors in [9] explored the influence of Human Resource Management Information Systems (HRMIS) on employee efficiency and satisfaction. Their findings highlight the significant impact of integrated information systems on enhancing organizational processes. This aligns with the current research, which aims to utilize Salesforce's integrated platform to streamline resource tracking and allocation, thereby improving overall project efficiency.

The study conducted by [10] discussed strategies for enhancing employee performance through digitalization in HR management. While their focus was on general HR practices, the principles of digital transformation and efficiency improvement through custom applications resonate with the development of a custom Salesforce application for resource management. This study emphasizes the potential of digital tools to optimize organizational processes, which is pertinent to our investigation into enhancing resource allocation and visibility.

The authors from the publication [11] investigated multi-project resource management methods suitable for research institutes. Their study emphasizes the importance of tailored resource management approaches that can adapt to varying project demands and organizational contexts. This resonates with Salesforce's customization capabilities, which allow for the development of tailored resource management solutions tailored to organizational needs. The study supports our approach of utilizing a custom Salesforce application to optimize resource allocation and improve project outcomes. The study referenced in [12] explored adaptive and priority-based resource allocation in mobile-edge computing. While their focus was on a different technological context, their insights into efficient resource utilization through adaptive algorithms provide relevant perspectives for optimizing resource allocation within organizational

settings. The principles of adaptive resource allocation can inform the development of intelligent resource management features within the Salesforce application, enhancing its effectiveness in dynamically allocating resources based on project needs.

Authors in [13] give an overview of the security challenges of CRM systems with an emphasis on data security. Their findings indicate that there are significant challenges when dealing with private data in CRM systems, especially since the number of cyberattacks has been increasing in recent years.

The primary goal of CRM systems is to foster customer loyalty, which is crucial for achieving high levels of customer retention. Without CRM systems, it would be challenging for enterprises to assess customer satisfaction, experience, and loyalty. Therefore, when setting up CRM systems and collecting customer data, it is essential to address users' concerns about their personal information and privacy. CRM systems are vital for gaining a competitive edge in the market as they enable enterprises to gather, process, visualize, share, and apply customer and market data, providing user-friendly and specific insights on key metrics [14]. Without an effective CRM system supported by robust data security measures, enterprises may struggle to maintain a competitive market position. As social media data breaches occur, customers are increasingly aware of the risks associated with data collection and storage, making them more conscious of potential cyber-attacks [15]. Customer data is vulnerable not only on social media and online platforms but also within enterprises, posing risks to data integrity [16]. Thus, enterprises must prioritize securing the customer data they collect to enhance CRM and build strong customer-to-business (C2B) relationships [17].

Organizations need advanced data storage systems to handle the vast, fast, and complex data from various touchpoints, often leading to non-relational databases like NoSQL. They must decide whether to use existing or acquire new resources and whether to handle data in-house or outsource it, always considering cost-benefit analyses. Understanding and meeting customer needs through CRM systems helps businesses build long-term relationships and retain customers, which is cheaper and more effective than acquiring new ones [18].

These studies collectively highlight the broader applicability of integrated information systems, such as Salesforce, in enhancing resource management practices across different organizational contexts. By leveraging Salesforce's robust platform and customization capabilities, organizations can potentially significantly improve resource utilization, project efficiency, and decision-making processes. Together, these studies provide a comprehensive framework for enhancing our resource management capabilities through the integration of HRMIS, the digitalization of HR functions, the customization of resource management approaches, and the implementation of adaptive algorithms. By applying these principles within the Salesforce platform, we can develop a robust system that enhances resource tracking, allocation, and overall organizational efficiency. Efficient resource management is crucial for organizational efficiency and project success. Traditional methods can often lead to inefficiencies and inaccuracies due to disparate systems lacking integration and real-time visibility. This research hypothesizes that a unified approach using Salesforce can improve these inefficiencies by centralizing resource management.

3 Methodology

In this section, we give an overview of the methodology used in this study. For this study, we performed qualitative analysis and, in the end, integrated the findings. First, we developed a custom Salesforce application using Apex code. The application is intended to track resource availability, bookings and project assignments. Additionally, the application includes UI enhancements such as color-coded visualizations and a search function.

After the application was developed, we performed a simulation to obtain simulated data that reflects the current organizational scenarios, including measurement of metrics such as resource utilization rates, project completion times, user satisfaction and security perception. Additionally, we have used surveys and interviews with users to perform qualitative analysis of the usage and their overall perception of the system.

The approach gives qualitative insights to offer a thorough understanding. For example, interviews can provide context on the reasons behind the system improvements, such as improved visibility of resource availability. This method allows for a comprehensive understanding of the process, allows cross-verifying of the results, and gives enough insights to give actionable recommendations.

3.1 Overview of Software Packages and Tools

The resource management solution was developed using advanced software tools and platforms, ensuring robust functionality and seamless integration within the Salesforce ecosystem:

- Salesforce Platform - Deployed as the cornerstone of the solution, Salesforce provided a scalable and secure environment for managing consultant data, automating workflows, and facilitating collaboration across teams. Its robust CRM capabilities were leveraged to streamline consultant lifecycle management from onboarding to project assignment and offboarding [19].
- Apex Programming Language - Crucial for implementing complex business logic and backend processes specific to consultant management. Apex facilitated the automation of critical workflows such as consultant assignment, skill matching, and performance tracking. Custom Apex triggers and classes were developed to enforce business rules, validate data integrity, and integrate with external systems seamlessly [20].
- Visualforce Pages and Lightning Components - Tailored user interfaces were created using Visualforce Pages and Lightning Components. These components offered a rich, responsive user experience by presenting data in intuitive layouts and enabling interactive features essential for managing consultant profiles, engagements, and performance metrics [21].
- Salesforce Lightning Framework - Employed to build dynamic and scalable applications within Salesforce. The framework ensured modern UI design principles, responsiveness across various devices, and enhanced usability through features like drag-and-drop functionalities and real-time updates [22].
- Salesforce Object Query Language (SOQL) and Salesforce Lightning Data Service - Utilized for efficient data retrieval and manipulation. SOQL queries were optimized

to fetch real-time information about consultant availability, engagement histories, and skill competencies. Lightning Data Service enhanced performance by caching data locally on users' devices, minimizing server calls for improved responsiveness [23].

- Integration with External Systems - Integrated seamlessly with external HR management systems and financial databases using Salesforce APIs. This integration enabled bi-directional data synchronization, ensuring that consultant records, payroll information, and project assignments remained up-to-date across platforms. Real-time data updates facilitated agile decision-making and accurate resource planning.

In this section, we have given an overview of the methodology and the software packages used for the implementation. In the following section we are going to present the initial findings.

4 Results

The deployment of the consultant management solution yielded substantial and quantifiable results, demonstrating its effectiveness in optimizing operational processes and enhancing organizational performance.

The following findings were obtained using the qualitative analysis of the data and the user responses regarding RQ1:

Streamlined Onboarding and Offboarding - Reduced the onboarding time for new consultants through automated workflows integrated with HR systems. Consultants were swiftly onboarded with pre-defined templates for contract creation, compliance checks, and access provisioning. These predefined templates for creating contracts, compliance checks, and access provisioning were consistent and efficient, with no avoidable delays and administrative overhead. For example, new consultants could start their assignments in just some days, which earlier took several weeks.

Enhanced Resource Allocation Efficiency - Improved resource utilization through real-time visibility into consultant availability and skill profiles. Automated alerts and notifications enabled resource managers to allocate consultants based on project demands and skill requirements promptly. This optimized the utilization of resources, while at the same time it saved a big chunk of time for resources from going into manual allocation processes.

Improved Decision-Making and Forecasting - Increased decision-making accuracy with centralized dashboards and real-time analytics. Predictive modeling and forecasting tools provided insights into future resource needs, enabling proactive planning and mitigating project risks. The single centralized dashboards and real-time analytics provided complete visibility to managers regarding the current resource allocation and future needs. As an example, managers would be able to know needs in a project months in advance and work on having the right type of consultants available on time.

User Adoption and Satisfaction - Achieved high user satisfaction with consultants and managers reporting improved productivity and ease of use. User training programs and continuous feedback loops ensured that the solution evolved to meet evolving business needs effectively.

Based on these findings, we can conclude that regarding RQ1, the answer is that the impact on resource utilization, availability, and skill set visibility from using custom Salesforce is very positive, and we can recommend custom Salesforce applications to increase resource utilization, availability, and visibility of skill sets within the application. A custom Salesforce application allows for the integration of automated workflows and real-time data analytics with all the associated advantages. By automating the onboarding and offboarding processes, much time and many resources were saved. In addition, this automation reduced the likely risk associated with human errors at every turn, ensuring that every compliance or contractual requirement was attended to. Having provided real-time insight into the availability and skills of consultants allowed for strategic and efficient allocation of resources. The higher visibility gave assurance that projects were staffed with the best group of consultants, thus improving project deliverables and client satisfaction. Predictive analytics helped improve the organization's decision-making ability by handling demands that were to be made in the future and to plan for resources accordingly. This proactive approach minimized the risks associated with a sudden demand for projects or unavailable consultants.

Regarding RQ2, the users' feedback underscored the tangible benefits and user-centric design of the consultant management solution and the usage of the UI elements, including the visualization and its ability to capture utilization complexities:

Ease of Use and Accessibility - Consultants appreciated the intuitive user interface that simplified task management, time tracking, and collaboration with project teams.

Drag-and-drop features and dashboards that were customisable to the various needs, added to self-explanatory navigation, which made this system even useful for those users who have little exposure to technology.

Real-Time Insights and Reporting - Managers commended the comprehensive dashboards that provided actionable insights into consultant performance metrics, project timelines, and budget utilization. This made it easy to track progress and notice bottlenecks in the process while making informed decisions to keep projects on course.

Time Savings and Efficiency Gains - Administrators noted a significant reduction in manual administrative tasks, allowing more time for strategic initiatives and client engagement activities.

Based on these findings, we can answer the RQ2 with yes. The UI elements used in custom applications are able to capture resource allocation complexities and the customers are very satisfied with the results. The UI elements of the custom Salesforce application were paramount to capturing the intricacies of resource allocation. Positive user reviews mirror the requirement of a user-centered design approach. Engaging users in the design process and iteration - which is continuous - gave them very high levels of user satisfaction and adoption rates, which ensured that the application would meet the evolving user needs. The visualization of complex data in a digestible format empowered managers and administrators to start making better and faster decisions, hence executing better project resources. The automation of routine tasks released a considerable amount of time for the administrator to engage in higher-order activities. This gain in efficiency translated into better resource management and improved client relations. The developed UI elements of this custom Salesforce application are very

effective at catching resource allocation complexities and increasing end-user satisfaction. The user interface development should be continued to deliver more functionality for retaining high levels of user engagement and productivity. It is the intuitive design and real-time reporting features of this solution that present a meaningful return on investment in resource management.

Regarding RQ3, the main security challenge in CRM software is data privacy protection. CRM software contains a lot of customer data and there are many different roles that need to have controlled access over the data. Authors in [13] identify that with the increased cybersecurity threats, there are higher risks in safeguarding customer data, especially data classified as personal. According to the GDPR rules, this data needs to be treated with utmost care and all risks need to be resolved urgently. Additionally, it requires a high level of transparency towards the users about the way the data is stored and used. The introduction of GDPR included an additional regulatory compliance requirement for CRM software. According to [24], Salesforce adheres to the GDPR regulatory requirements. Regarding the other cybersecurity risks, we analyse the security architecture of the Salesforce application, which has the following features: Data Encryption - Employed Salesforce's native encryption features to secure sensitive data at rest and in transit, adhering to industry standards such as GDPR and HIPAA.

Role-Based Access Control (RBAC) - Implemented granular access controls using Salesforce profiles, permission sets, and field-level security to restrict data access based on user roles and responsibilities. For instance, a consultant may view all his own data regarding the projects he handles and his working time but may not view the financial data or data of another consultant.

Compliance Measures - Conducted regular security audits and implemented robust access logging mechanisms to monitor user activities and ensure compliance with internal policies and regulatory requirements.

Secure API Integrations - Implemented OAuth authentication and secure API endpoints to facilitate safe and secure integration with external systems, preventing unauthorized data access and ensuring data integrity.

Additionally, we analyzed the user's perception of security with the usage of custom-developed applications. The results show that regarding RQ3, the Salesforce CRM is highly effective secure and complies with the latest cybersecurity standards and requirements. Additionally, as CRM, it allows tightly secured management of user access, data access, and data security.

4.1 Survey Analysis

The survey aimed to assess the utility and effectiveness of CRM systems, focusing on Salesforce, in managing company resources. The survey comprised 68 questions, covering various aspects of CRM usage, ease of use, integration with other tools, impact on productivity, personalization options, cost-effectiveness, data security, customer satisfaction, and specific features offered by different CRM systems, including Salesforce, Zendesk, Microsoft Dynamics 365, HubSpot, and Zoho, as well as questions on the usefulness of the custom resource management app that was made.

Most respondents reported using a CRM system, with Salesforce being the most frequently cited. This prevalence underscores Salesforce's dominance in the CRM market and its critical role in resource management for many companies. The survey revealed that users generally recognize Salesforce as a powerful tool, particularly in managing resources effectively. However, the feedback also highlighted several areas for potential improvement.

In terms of ease of use, responses were varied. While many users appreciate the comprehensive functionalities that Salesforce offers, some found the system's interface and navigation to be less intuitive. This suggests that while Salesforce provides robust features, the complexity of its interface may hinder user experience, particularly for those who are not as technically proficient.

Integration with other tools was another critical area explored in the survey. Many respondents expressed satisfaction with Salesforce's ability to integrate with a wide range of other business applications. This capability is crucial for organizations seeking to streamline their operations and ensure seamless data flow across different platforms. Despite this, some users noted challenges in achieving full integration, particularly with non-Salesforce tools, indicating that there might be room for improvement in making integrations more user-friendly and comprehensive.

The impact of Salesforce on productivity was overwhelmingly positive. A significant portion of respondents agreed that Salesforce contributes to increased efficiency and effectiveness in managing company resources. This is particularly true in the areas of sales automation and resource allocation, where Salesforce's advanced features help teams to optimize their workflows and make more informed decisions. Automating repetitive tasks and managing complex resource allocations was frequently highlighted as a key benefit.

Personalization of the CRM system was another area of interest. Most respondents indicated that the ability to customize Salesforce to meet their specific business needs was highly important. Salesforce's flexibility in this regard was generally praised, although some users felt that certain customizations could be made more accessible or easier to implement without requiring extensive technical expertise. When it came to cost-effectiveness, opinions were divided. Some respondents felt that Salesforce, while expensive, offered sufficient value to justify its cost. However, others expressed concerns that the pricing structure might be prohibitive, particularly for smaller organizations. This suggests a perception that while Salesforce is a premium product, its cost may not always align with the value perceived by all users.

Data security, a critical concern for any CRM system, was generally rated highly by respondents. Salesforce was seen as a reliable platform in terms of protecting sensitive business and customer data. However, a small number of respondents did express some concerns, indicating that while Salesforce's security features are robust, ongoing vigilance and improvements are necessary to maintain trust. The survey also explored specific features within Salesforce that are crucial for effective resource management. The color-coded display system used in Salesforce for resource allocation was particularly noted. This feature, which visually represents resource utilization, was appreciated for its ability to quickly convey critical information. However, some respondents indicated a need for greater precision and clarity in these visualizations,

suggesting that while the feature is useful, it could be further refined to improve its effectiveness.

In addition to resource management, respondents were asked about the overall impact of Salesforce on their operations and competitiveness. The majority were optimistic, with many believing that Salesforce significantly enhances their company's efficiency and competitiveness in the market. This positive outlook highlights the strategic importance of Salesforce in helping organizations manage their resources more effectively, reduce operational costs, and improve decision-making processes.

The analysis of the survey responses to the resource management app reveals several key insights that are essential for understanding user satisfaction and the app's potential impact on business operations. While the feedback is generally positive, the analysis will maintain a critical perspective, ensuring that the app's strengths are highlighted without ignoring areas where improvements could be beneficial.

The ease of navigation and clarity of the user interface are crucial factors in user satisfaction. Respondents largely found the app easy to navigate, which indicates a well-designed interface. However, achieving a balance between simplicity and functionality is critical. While users rated the clarity of the interface highly, there is always room for enhancing the intuitiveness of complex features to accommodate users with varying levels of technical expertise. One of the app's standout features is its ability to search employees based on certifications, skills, and other attributes. The respondents rated This functionality as highly useful, emphasizing the app's value in efficiently managing human resources. Quickly locating and assessing employee capabilities is essential for optimizing resource allocation, particularly in dynamic work environments.

Similarly, the management of different types of employee certifications, including new, renewed, and expired ones, received positive feedback. This aspect is vital for companies that need to ensure compliance with industry standards and regulations. The app's ability to provide clear and accessible views of certification data helps managers maintain up-to-date records, which can prevent lapses that might otherwise lead to operational disruptions or penalties. The engagement management interface, particularly its intuitiveness, was generally well-received, although this is an area where user feedback suggests there could be incremental improvements. A system that manages revenue-generating engagements alongside non-revenue activities such as sick leave and vacations must balance comprehensiveness with simplicity to prevent user overwhelm.

The app's reporting capabilities are another critical feature, with users appreciating the clarity and utility of the report generation component. However, ensuring that the instructions for data entry and report generation are clear and accessible remains a priority. The ability to filter resources in various ways when creating reports is highly valued, and this feature is seen as contributing significantly to more informed decision-making.

Adaptability is a key strength of the app, with users expressing satisfaction with how easily the app can be tailored to their specific needs. This adaptability is crucial in a business environment where processes and requirements can change rapidly. However, maintaining this flexibility while ensuring that the app remains user-friendly is a balancing act that requires ongoing attention.

Regarding visual data representation, users found the color-coded display particularly useful for identifying underutilized resources and making decisions about resource allocation. The precision of this feature in reflecting true resource utilization levels was generally rated positively, although continuous refinement is necessary to ensure accuracy and reliability.

Finally, the impact of the app on business operations was perceived positively, with users expecting it to enhance productivity, reduce operational costs, and improve overall efficiency. The ability of the app to track non-revenue-generating activities and team dynamics (such as new hires or departures) was also seen as effective, further underlining its potential as a comprehensive tool for resource management.

In conclusion, the survey findings illustrate that while Salesforce is highly valued for its comprehensive features and ability to manage resources effectively, there are areas where users see opportunities for improvement. The complexity of the interface, challenges with integration, and concerns about cost highlight the need for Salesforce to continue evolving to meet the diverse needs of its user base. Nevertheless, the overall impact of Salesforce on productivity and competitiveness is seen as overwhelmingly positive, reaffirming its position as a leading CRM solution in the market. This analysis underscores the importance of ongoing innovation and user-focused improvements to maintain Salesforce's competitive edge in an increasingly crowded market.

While the resource management app received favorable feedback in several key areas, including navigation, certification management, reporting, and adaptability, it is crucial to continue refining these features to meet the evolving needs of users. The app's success will ultimately depend on its ability to maintain a high level of user satisfaction while continuously improving its functionality and usability. This analysis provides a balanced view, acknowledging the app's strengths while also identifying areas where further development could enhance its value to businesses.

5 Conclusion

The consultant management solution made a big difference in how the company operates. It made hiring and letting go of consultants much faster, cutting the time from weeks to just days. It also helped managers see who was available and what skills they had in real time, making it easier to assign the right people to projects. The dashboards and analytics gave managers a clear view of what was happening and what would be needed in the future, helping them plan better and avoid problems. Everyone found the system easy to use, which made them more productive and happy with their work. It also saved administrators a lot of time by cutting down on manual tasks, allowing them to focus on more important things.

On the security side, the Salesforce CRM did a great job of keeping customer data safe. It followed all the important data privacy rules, like GDPR, by using encryption and strict access controls to make sure only the right people could see the data. Secure API integrations also helped protect the information. Regular security checks and logging of user activities made sure everything stayed safe and compliant. Users were very satisfied with these security measures, showing that the solution was effective at protecting data. Overall, the custom Salesforce application brought big improvements

in operations, resource management, user satisfaction, and data security, making it a great choice for companies.

This study highlights Salesforce's role in improving resource management through a custom application. By evaluating its impact, the research demonstrates how Salesforce enhances resource allocation, visibility into availability and skills, and overall project efficiency. Utilizing Salesforce's features and customization options enables organizations to optimize resource use and make informed decisions. This approach enhances operational efficiency and supports better project outcomes.

Moving forward, leveraging Salesforce can significantly streamline processes and promote sustainable growth. Further research and implementation will refine these strategies, maximizing Salesforce's impact across industries.

Acknowledgements

The work presented in this paper is partially funded by the Ss Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering.

References

[1] What is Salesforce? Last accessed 05.07.2024. https://www.salesforce.com/products/what-is-salesforce/

[2] Salesforce. Last accessed 05.07.2024. https://www.techtarget.com/searchcustomerexperience/definition/Salesforcecom

[3] Why is Salesforce CRM Better than Other CRMs? A List of Features Every Business Should Know. Last accessed 05.07.2024. https://www.linkedin.com/pulse/why-salesforce-crm-better-than-other-crms/

[4] Five Common Salesforce Security Risks. Last accessed 05.07.2024. https://www.flosum.com/blog/five-common-salesforce-security-risks

[5] The Biggest Security Risks to Your Salesforce Org. Last accessed 05.07.2024. https://www.varonis.com/blog/security-risks-to-your-salesforce-org

[6] Jahan, S., Jahan, S., Ahmad, F.: Ensuring data security on salesforce: A comprehensive review of security measures and best practices. International Journal of Engineering and Management Research **13**(2) (2023)

[7] Data Privacy and Security in Salesforce. Last accessed 05.07.2024. https://www.flosum.com/blog/data-privacy-and-security-in-salesforce-how-enterprises-can-ensure-customer-confidence

[8] Oril: 5 Strategies for Strengthening Salesforce Security. Last accessed 05.07.2024. https://medium.com/@oril_/5-strategies-for-strengthening-salesforce-security-bbb15af2c7a4

[9] Zahari, A.M., *et al.*: The influence of human resource management information system (hrmis) application towards employees efficiency and satisfaction. Journal of Physics: Conference Series **1019**, 012077 (2018)

[10] Tommy, R., Kurniawan, C., Niccosan, Makalew, B.A.: Improving employee performance through digitalization: Designing a web based human resource management. In: 2022 International Conference on Information Management and Technology (ICIMTech), pp. 655–660 (2022). IEEE

[11] Li, X.B., Nie, M., Yang, G.H., Wang, X.: The study of multi-project resource management method suitable for research institutes from application perspective. Procedia Engineering **187**, 191–197 (2017)

[12] Sharif, Z., Jung, L.T., Razzak, I., Alazab, M.: Adaptive and priority-based resource allocation for efficient resources utilization in mobile-edge computing. IEEE Internet of Things Journal **10**(4), 3079–3093 (2023)

[13] Bakator, M., orević, D., Ćoćkalo, D., Ćeha, M., Bogetić, S.: Crm and customer data: Challenges of conducting business in digital economy. Journal of Engineering Management and Competitiveness (JEMC) **11**(2), 85–95 (2021)

[14] Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Ravald, A., Sarantopoulos, P., Villarroel Ordenes, F., Zaki, M.: Customer experience management in the age of big data analytics: A strategic framework. Journal of Business Research **116**, 356–365 (2020) https://doi.org/10.1016/j.jbusres.2020.01.022

[15] Hu, T., Wang, K.-Y., Chih, W., Yang, X.-H.: Trade off cybersecurity concerns for co-created value. Journal of Computer Information Systems, 1–16 (2018) https://doi.org/10.1080/08874417.2018.1538708

[16] Powell, M.: 11 Eye Opening Cyber Security Statistics for 2019. Last accessed 05.07.2024. https://www.cpomagazine.com/tech/11-eye-opening-cyber-security-statistics-for-2019/

[17] Bakator, M., orević, D., Ćoćkalo, D., Ćeha, M., Bogetić, S.: Crm and customer data: Challenges of conducting business in digital economy. Journal of Engineering Management and Competitiveness **11**(2), 85–95 (2021) https://doi.org/10.5937/jemc2102085B

[18] Optimizing CRM: A Framework for Enhancing Profitability and Increasing Lifetime Value of Customers. Last accessed 05.07.2024. https://www.mmaglobal.org/_files/ugd/3968ca_dbde4f639d944242b91d677bf3fd5461.pdf#page=147

[19] Salesforce platform. Last accessed 01.07.2024. https://www.salesforce.com/eu/products/salesforce-platform/

[20] What is Apex? Last accessed 01.07.2024. https://developer.salesforce.com/docs/

atlas.en-us.apexcode.meta/apexcode/apex_intro_what_is_apex.htm

[21] Use Lighting Components and Visualforce Pages. Last accessed 01.07.2024. https://developer.salesforce.com/docs/atlas.en-us.lightning.meta/lightning/components_visualforce.htm

[22] Lighting Components Framework. Last accessed 01.07.2024. https://help.salesforce.com/s/articleView?id=sf.aura_overview.htm&type=5

[23] Salesforce Object Query Language (SOQL). Last accessed 01.07.2024. https://developer.salesforce.com/docs/atlas.en-us.soql_sosl.meta/soql_sosl/sforce_api_calls_soql.htm

[24] Compliance engineered for the Cloud. Last accessed 01.07.2024. https://compliance.salesforce.com/en/gdpr

Comparative Analysis of National E-Health Initiatives: Development Trajectories in Croatia, Slovenia, and North Macedonia

Zhaklina Chagoroska[1] and Smilka Janeska Sarkanjac[1]

[1]Ss Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Rugjer Boshkovich 16, Skopje, 1000, North Macedonia.

Contributing authors: zakichagoroska@gmail.com; smilka.janeska.sarkanjac@finki.ukim.mk;

Abstract

This paper offers a comparative analysis of the national e-health initiatives in Croatia, Slovenia, and North Macedonia. The stages of digitization in these countries vary significantly, shaped by their unique historical, economic, and policy contexts. This study summarizes key advancements, strategies, and outcomes in each country, highlighting the pivotal role of e-health in modern healthcare systems. E-health initiatives are essential for enhancing service delivery, patient outcomes, and cost efficiency.

By examining the historical background, design processes, implementation mechanisms, and outcomes, this paper provides a comprehensive overview of each country's journey in digitizing healthcare services. The research is grounded in detailed case studies, offering insights into broader regional efforts and the contextual differences in e-health development.

Social mechanism theory is employed to evaluate and explain how and why each e-health system in the analyzed countries achieves its intended results. This evaluation approach, based on the work of Melloni, Pesce, and Vasilescu [1], provides a process-based framework for understanding the relationship between processes, contexts, and outcomes in e-health initiatives.

Keywords: E-health initiatives, Healthcare digitization, Comparative analysis

1 Introduction

The digitization of healthcare systems, commonly referred to as e-health, has become a pivotal aspect of modernizing national health services worldwide. E-health initiatives aim to enhance service delivery, improve patient outcomes, and increase cost efficiency through the integration of information technology into healthcare processes. This study provides a comparative analysis of national e-health initiatives in Croatia, Slovenia, and North Macedonia, each at different stages of digital healthcare development influenced by distinct historical, economic, and policy contexts.

Slovenia, Croatia, and North Macedonia, all former Yugoslav republics, share a common historical background but have embarked on divergent paths in their healthcare digitization efforts post-independence. Slovenia, an EU member since 2004, and Croatia, since 2013, have advanced significantly in their e-health initiatives, benefiting from EU funding and strategic planning. Conversely, North Macedonia, an EU candidate, has faced more substantial challenges but has also made notable strides in digitizing its healthcare system.

This paper examines the development trajectories of e-health initiatives in these three countries, providing detailed case studies to understand their unique approaches, challenges, and successes. By analyzing the historical background, design processes, implementation mechanisms, and outcomes of e-health initiatives, this study offers a comprehensive overview of how each country has navigated the complexities of digitizing healthcare services.

Using social mechanism theory, the research evaluates why and how each e-health system in these countries contributes to achieving the intended outcomes. This approach, grounded in the work of Melloni, Pesce, and Vasilescu (2016), provides a process-based framework for understanding the relationship between context, process, and outcomes in e-health development. The findings aim to offer insights and recommendations that could guide future e-health initiatives in similar contexts, particularly focusing on the transformative potential of digital health in enhancing healthcare quality and accessibility.

Through this comparative analysis, the study seeks to highlight the critical factors driving successful e-health implementation and the persistent challenges that need addressing. The ultimate goal is to contribute to the ongoing discourse on digital health transformation and to provide actionable insights for policymakers, healthcare providers, and other stakeholders involved in e-health initiatives.

2 Historical Context of Healthcare System Digitalization

2.1 Slovenia

Before its independence in 1991, Slovenia was part of Yugoslavia, where healthcare was centrally planned and publicly funded. The healthcare system was characterized by state ownership of health facilities and the employment of healthcare professionals by the state. The foundations of Slovenia's healthcare system were laid during this period,

emphasizing universal healthcare coverage, public health initiatives, and a network of primary, secondary, and tertiary healthcare institutions.

After declaring independence from Yugoslavia in 1991, Slovenia began reforming its healthcare system. The transition involved decentralization and a shift towards a mixed public-private model. The Health Insurance Institute of Slovenia (HIIS), established in 1992, became the cornerstone of the new healthcare financing system, collecting mandatory health insurance contributions from employers and employees to fund healthcare services.

The early 2000s marked the initial steps towards integrating information technology into the healthcare system. Efforts focused on creating databases and electronic health records (EHRs). The government developed a strategic framework for eHealth, recognizing the potential of digitalization to improve healthcare delivery and efficiency. Development of a National Health Information System (NHIS) began, aiming to connect healthcare providers and streamline data exchange. Several pilot projects were launched to test and implement eHealth solutions, such as electronic prescriptions (e-Prescriptions) and telemedicine services. As an EU member since 2004, Slovenia benefited from EU funding and expertise, accelerating the adoption of digital health technologies.

The Ministry of Health launched the eHealth Program in the 2010s, which outlined key objectives for digitalizing the healthcare system. The program aimed to enhance interoperability, data security, and patient access to health information. In 2010, the eZdravje (eHealth) portal was introduced, providing citizens with online access to their health records, appointments, and other healthcare services. Efforts were made to establish an interoperability framework to ensure seamless data exchange between different healthcare providers and systems.

Enacted in 2014, the Health Information Act provided a legal basis for the collection, processing, and exchange of health data. It aimed to protect patient privacy and ensure data security. The Ministry of Health developed a strategic plan for 2020-2025, focusing on expanding digital health services, improving infrastructure, and enhancing the quality of care through technology.

The COVID-19 pandemic in the 2020s accelerated the adoption of digital health solutions. Remote consultations, telehealth services, and digital contact tracing tools became essential components of the healthcare response. Slovenia continues to invest in eHealth infrastructure, including the expansion of the National Health Information System and the integration of new technologies like artificial intelligence and big data analytics. Efforts are ongoing to enhance patient-centric services, such as personalized health records, mobile health applications, and online portals for managing healthcare appointments and accessing medical information.

Slovenia's journey towards healthcare system digitalization has been marked by significant milestones, from its early steps in the post-independence era to the comprehensive eHealth initiatives of the 2010s and beyond. The combination of strategic planning, legislative support, and EU collaboration has positioned Slovenia as a leader in digital health within the region. Continued investment in infrastructure and technology aims to further improve the efficiency, accessibility, and quality of healthcare services for its citizens.

2.2 Croatia

Before declaring independence in 1991, Croatia was part of Yugoslavia, where the healthcare system operated under centralized planning, providing universal healthcare coverage and state-owned healthcare facilities. This system laid a strong foundation for public health and primary care services, which Croatia inherited upon independence. Emphasis was placed on preventive care, supported by an extensive network of primary healthcare centers alongside secondary and tertiary healthcare institutions.

Following independence in 1991, Croatia underwent significant political and economic changes, including reforms in the healthcare sector. The Croatian Health Insurance Fund (HZZO) was established in 1993, marking a pivotal shift towards a new era in healthcare financing. The digitization process commenced in 1994 with the introduction of health insurance cards equipped with magnetic stripes, setting the stage for a more integrated healthcare information system.

By 1998, hospitals began issuing personal invoices for hospital care on 3.5" floppy disks to the HZZO, signaling a crucial step towards digitalizing healthcare operations. The implementation of the health insurance information system ZOROH was pivotal in managing the registry of health-insured persons.

The e-Croatia initiative launched in 2001 aimed at connecting healthcare providers, the Croatian Health Insurance Fund, and public health institutes. Central to this initiative was the development of the Central Health Information System of the Republic of Croatia (CEZIH). The National e-Health project, initiated in 2003 with financing from a World Bank loan, further advanced the development of a central eHealth system. In 2007, smart cards for health professionals were introduced to enhance security and efficiency in health data exchange through secure authentication and digital signatures.

The 2010s saw significant expansions in Croatia's eHealth capabilities. By 2011, nationwide implementation of e-Prescriptions and e-Referrals significantly improved the efficiency of healthcare services. From 2012 onwards, functionalities such as e-Booking and e-Waiting lists were developed to streamline the scheduling of health services. The introduction of the e-Citizens portal in 2014 provided citizens with access to various eHealth services, including health records, prescriptions, and appointments, with a mobile version introduced in 2017.

Strategic planning and legislative support have been integral to Croatia's eHealth development. The Ministry of Health issued the "Strategic Plan of eHealth Development in the Republic of Croatia" in 2017, outlining the country's future eHealth direction. The Health Data and Information Act of 2019 strengthened personal data protection in healthcare and established an eHealth authority for managing health data governance. Croatia's medium-term eHealth strategic framework for 2020-2027, supported by the EU Commission, emphasizes specific eHealth activities and international best practices.

The COVID-19 pandemic in the 2020s accelerated the adoption of digital health solutions in Croatia, including electronic referrals and remote consultations. Despite notable progress, challenges persist, such as the shortage of IT personnel in healthcare institutions and the need for enhanced interoperability of IT systems and further development of electronic health records. Current efforts are focused on

advancing comprehensive electronic health records, centralized scheduling of specialist consultations, and continuous enhancement of the eHealth infrastructure.

In summary, Croatia's path towards healthcare system digitalization has been characterized by strategic reforms and initiatives aimed at integrating information technology into healthcare delivery. Beginning with foundational reforms in the early 1990s and culminating in the establishment of robust eHealth infrastructure and services, Croatia continues to invest in and prioritize the development of its eHealth ecosystem to improve the quality, accessibility, and efficiency of healthcare services.

2.3 North Macedonia

Before gaining independence in 1991, North Macedonia was part of Yugoslavia, inheriting a centrally planned healthcare system characterized by universal coverage and state ownership. This system prioritized preventive care and primary healthcare services, establishing a robust network of healthcare institutions that persisted post-independence.

Following independence, North Macedonia underwent significant political, economic, and social transformations, necessitating reforms in its healthcare system. Decentralization efforts began to reshape healthcare delivery, alongside the establishment of new institutions and health insurance schemes. The Ministry of Health assumed a pivotal role in regulating healthcare policies and standards during this period.

In the late 1990s and early 2000s, digital initiatives in North Macedonia's healthcare sector were initiated, spearheaded by the Health Insurance Fund (HIF). These initiatives aimed to modernize healthcare management and administration. The privatization of primary healthcare general practitioners (GPs) between 2001 and 2006 coincided with early efforts in computerizing healthcare operations, introducing basic electronic tools and systems.

A significant milestone occurred in 2006 with the adoption of the Strategy on the Development of an Integrated Health Information System (IHIS). This strategy laid out the legal, scientific, organizational, and functional requirements for a comprehensive health information system, setting the stage for further digitization efforts. By 2009, the informatization of the Health Insurance Fund enabled healthcare institutions to electronically submit results, reports, and financial documents.

The pivotal step towards digital healthcare came in the early 2010s with the introduction of the National System for Electronic Records, known as My Term (Moj Termin). This integrated system provided modules for electronic scheduling, patient referrals, electronic health records, and an e-health portal for citizens. In 2015, the eHealth Directorate was established within the Ministry of Health to manage and develop the national health information system, institutionalizing eHealth initiatives.

Legislative support played a crucial role in advancing eHealth capabilities. The adoption of the Law on Health Records in 2009 regulated electronic and paper-based health records, establishing the framework for the National System for Electronic Health Records managed by the Ministry of Health. Amendments to the Health Care Law in 2015 further clarified segments of the integrated national health information system, providing a robust legal structure for ongoing eHealth developments.

The COVID-19 pandemic underscored the importance of robust digital health systems, accelerating the adoption of electronic health solutions such as e-prescriptions and remote consultations. Despite progress, North Macedonia faces challenges in eHealth, including the need for stable information infrastructure, data standardization for interoperability, comprehensive eHealth strategy development, and improving IT literacy among citizens and healthcare professionals.

In summary, North Macedonia has made steady strides in healthcare system digitalization since the early 2000s. With foundational legislative support, the implementation of the My Term system, and the establishment of the eHealth Directorate, the country is committed to enhancing its digital health infrastructure. Ongoing efforts focus on improving interoperability, IT literacy, and strategic planning to ensure a comprehensive and efficient healthcare system that meets the needs of its citizens.

3 Comparative Insights

Upon thorough examination of e-health projects in Croatia, Slovenia, and North Macedonia, certain trends and points of comparison become apparent, as illustrated in Table 1. This table provides a clear overview of the trends and points of comparison between the e-health initiatives in the three countries. It highlights both the similarities and differences, as well as the progress and challenges faced by each nation.

Slovenia and Croatia are more advanced in their health digitization efforts compared to North Macedonia, largely due to EU membership and access to funding and resources. Both have established comprehensive eHealth strategies and systems that integrate health data and services effectively. Both Croatia and Slovenia have made significant strides in the digitization of their healthcare systems, with Slovenia showing a slightly more advanced integration and use of e-health solutions. The success in both countries has been driven by strong government initiatives, international support, and strategic planning. However, both face ongoing challenges that need addressing to further enhance their e-health landscapes, such as better integration of IT systems in Croatia and the development of a long-term digital health strategy in Slovenia. North Macedonia has made significant progress with initiatives like Moj Termin System but faces ongoing challenges in infrastructure, standardization, and strategic planning. It is in an earlier stage of digitization compared to its EU counterparts.

4 Social Mechanisms and e-health: Learning from the examples of Slovenia and Croatia

Social mechanism theory, as applied in this study, focuses on identifying and explaining the processes through which certain outcomes are produced in specific contexts. The theory helps bridge the gap between the initial conditions (such as historical, economic, and policy contexts) and the outcomes of e-health initiatives (like improved healthcare delivery and patient outcomes). Social mechanism theory in this study contributes to:

Understanding Contextual Differences - The social mechanism theory allows the researchers to delve into how the unique historical and socio-political contexts of Slovenia, Croatia, and North Macedonia have shaped their e-health initiatives. By

Table 1 Overview of Healthcare Systems in Slovenia, Croatia, and North Macedonia (Part 1)

Category	Slovenia	Croatia	North Macedonia
Context and Overview [2] [3] [4] [5] [6] [7] [8] [9] [10]	• Population: 2,107,007 • Health Expenditure: 10.1 % of GDP • Aging Population: 21% over 65 years • EU Member since 2004 • Healthcare system: Family medicine model, primarily public ownership, mandatory health insurance	• Population: 3,888,529 • Health Expenditure: 7.4% of GDP • Aging Population: 20.7% over 65 years • EU Member since 2013 • Healthcare system: Community-oriented primary care model, primarily public ownership, mandatory health insurance	• Population: 1,836,713 • Health Expenditure: 7.25% of GDP • Aging Population: 14.48% over 65 years • EU Candidate • Healthcare system: General practitioners (GPs) providing primary care, primarily public ownership, mandatory health insurance
Process Design Features [3] [11**?**] [12] [4] [13] [14] [15] [16] [17] [18]	• Early Initiatives: Began with automatic processing of prescriptions in 1974, evolving into a national medicines information system. • eHealth Strategy: Launched in 2005 with goals for a fully integrated national information system by 2023. • e-Zdravje Project: Major digitalization project from 2008-2015, including ePrescriptions, eAppointments, and the zVEM portal.	• Early Digitization: Initiated in the late 1990s with the development of health information systems. • eHealth Strategy: Comprehensive eHealth strategy developed in the 2000s, with a focus on electronic health records and telemedicine. • CEZIH: Central Health Information System of the Republic of Croatia (CEZIH) established to integrate health data across the country, including ePrescriptions, eReferrals, and patient portals.	• Early Computerization: Intensive computerization of primary healthcare from 2001 to 2006. • Integrated Health Information System Strategy: Adopted in 2006 to outline the development of a national health information system. • My Term (Moj Termin): Launched as a national system for electronic records, collecting data across all levels of healthcare since 2013.

Table 2 Overview of Healthcare Systems in Slovenia, Croatia, and North Macedonia (Part 2)

Category	Slovenia	Croatia	North Macedonia
Mechanisms and Promotion [19] [11] [20] [21] [] [22] [10] [23] [24] [25] [26]	• Institutional Leadership: Managed by the Ministry of Health with significant roles played by the National Institute of Public Health (NIPH) and Health Insurance Institute of Slovenia (HIIS). • Legislation: Various laws and strategies underpin the digitization efforts, including the Resolution on the National Healthcare Plan 2016–2025. • Funding: EU co-financed projects like e-Zdravje.	• Institutional Leadership: Managed by the Ministry of Health with the support of the Croatian Health Insurance Fund (HZZO). • Legislation: Health Care Act and related regulations support the digitization of healthcare services. • Funding: EU funding and national resources used to develop and maintain eHealth infrastructure.	• Government Coordination: The Ministry of Health coordinates with the Health Insurance Fund and the Directorate for E-Health. • Legislation: Law on Health Records (2009) and amendments to the Healthcare Law (2015) support electronic records and data processing. • Funding: Initial support through a World Bank loan, with continuous efforts by the Ministry of Health.
Outcomes and Challenges [20] [27] [28] [29] [30] [31] [32] [33] [34]	• eHealth Solutions: Successful implementation of ePrescriptions (96% of all prescriptions), eAppointments, and the zVEM portal with millions of visits. • Data Integration: Central Register of Patient Data (CRPD) serves as a core eHealth database, facilitating significant data transactions. • Recognition: Ranked high in the EU Digital Economy and Society Index, with notable savings and efficiency gains in the health system.	• CEZIH: Fully operational, integrating health data across all levels of care. High adoption rate of ePrescriptions and eReferrals. • eHealth Services: Patient portals, telemedicine services, and electronic health records improve accessibility and continuity of care. • EU Integration: Participation in EU health data initiatives and networks, enhancing interoperability and standards.	• Moj Termin System: Includes electronic scheduling, health records, e-prescriptions, immunization records, and various administrative modules. • Challenges: Need for stable infrastructure, data standardization, comprehensive eHealth strategy, IT literacy, and continuous training for healthcare workers. • Partial Integration: Systems from various institutions are integrated to some extent, but full data exchange and utilization are still limited.

doing so, it helps in identifying why certain strategies succeeded in one country but faced challenges in another. For instance, Slovenia's advanced e-health system is partly attributed to its strong institutional leadership and EU support, while North Macedonia's slower progress is linked to infrastructure challenges and the need for a comprehensive eHealth strategy.

Identifying Key Processes and Interventions - The theory emphasizes the identification of the underlying processes or mechanisms that lead to successful e-health implementation. In Slovenia and Croatia, mechanisms such as strong governmental initiatives, international support, and strategic planning have been crucial. Recognizing these processes informs the recommendations for North Macedonia, suggesting similar interventions like promoting a culture of acceptance, flexible regulation, and learning from more developed systems.

Highlighting the Role of Stakeholders - Social mechanisms involve not just structural elements but also the actions and interactions of various stakeholders, including government bodies, healthcare providers, IT professionals, and patients. By applying this theory, the paper can propose recommendations that focus on stakeholder engagement, such as organizing national campaigns to raise awareness, involving communities in e-health initiatives, and improving IT literacy among healthcare professionals.

The findings indicate that more developed countries such as Slovenia and Croatia, with well-established e-health systems, focus on promoting and encouraging initiatives rather than strict social mechanisms. This can be attributed to a culture of embracing innovation and technology. In the following are the recommendations resulting from this analysis, which refer to improvement in the development and, above all, the acceptance of e-health by all stakeholders. The application of these recommendations can help transform the e-health system in the Republic of North Macedonia, enabling easier acceptance and integration of new technologies and improving the quality of health services:

4.1 Promotion and encouragement

By organizing national campaigns to raise awareness of the benefits of e-health, using various media and platforms to reach all segments of society. Regular organization of information sessions and workshops for health professionals and patients will significantly act in the direction of promoting the use of e-health technologies.

4.2 Culture of acceptance

Introducing e-health modules in medical and health education programs will encourage the acceptance of new technologies from the beginning of the career. Also, active involvement of the community and patients in the development and implementation of e-health initiatives will create a sense of ownership and acceptance.

4.3 Flexibility in regulation

Developing and adapting existing regulation, in order to be flexible enough to adapt to rapid changes in technology, but also strict enough to ensure security and data protection. It is necessary to aim for the gradual introduction of regulations and

standards, allowing time for adaptation and training of employees in health facilities as well as users of health services.

4.4 Learning from the examples of the more developed

Increased cooperation with countries like Slovenia and Croatia, through the exchange of knowledge, experiences and best practices in the implementation of e-health systems. Adapting successful models and strategies from these countries to the local context, taking into account specific social and cultural conditions.

5 Recommendations for improving e-health in North Macedonia

Based on the comparative analysis of e-health initiatives in Slovenia, Croatia, and North Macedonia, several recommendations can be drawn for North Macedonia to further develop and improve its digital healthcare system:

5.1 Develop a Comprehensive eHealth/Digital Health Strategy

Creation of a detailed eHealth/Digital Health strategy, which will include clear goals, timelines and responsibilities. This strategy should cover all aspects of healthcare digitization, including infrastructure, interoperability, data security and user training. The involvement of all stakeholders, including government bodies, healthcare providers, IT professionals and patients, in the development and implementation of the strategy is essential.

5.2 Investing in infrastructure

Investment in upgrading and maintaining a stable and scalable IT infrastructure is required to support the growing demands of eHealth services. This includes high speed internet connection, secure servers and reliable data storage solutions. Establish and implement national standards for full data exchange and interoperability to ensure seamless integration between different healthcare providers and systems.

5.3 Improving data security and privacy

There is a need to strengthen the legislative framework to protect the privacy of patient data and ensure compliance with international standards, such as GDPR. Implementing strong cybersecurity measures to protect sensitive health data from breaches and cyber threats should be a primary task.

5.4 Promote IT Literacy and Training

Provide continuous training and support for healthcare professionals to enhance their IT literacy and competence in using digital health tools. Conduct public awareness campaigns to educate citizens about the benefits of eHealth and how to use digital health services effectively.

5.5 Leverage International Support and Collaboration

Seek funding and technical assistance from the European Union and other international organizations to support eHealth initiatives. Collaborate with neighbouring countries and participate in international eHealth projects to learn from best practices and innovative solutions.

5.6 Expand and Integrate Digital Health Services

Continue the development and implementation of comprehensive EHRs that are accessible across all levels of care. Expand telemedicine services and remote patient monitoring to increase healthcare access, especially in rural areas. Enhance the functionality of patient portals to provide easier access to health records, appointment scheduling, and health information.

5.7 Monitor and Evaluate Progress

Establish key performance indicators (KPIs) to monitor the progress of eHealth initiatives and measure their impact on healthcare delivery and patient outcomes. Regularly review and update eHealth strategies and policies based on performance data, feedback from stakeholders, and technological advancements.

5.8 Address Specific Challenges

Address the shortage of IT personnel in healthcare by offering incentives and training programs to attract and retain skilled professionals. Focus on data standardization to improve interoperability and data exchange across different healthcare systems and platforms. By focusing on these recommendations, North Macedonia can build a more robust, efficient, and patient-centered digital healthcare system, drawing on the experiences and successes of Slovenia and Croatia while addressing its unique challenges.

6 Conclusion

The digitization of healthcare in Slovenia, Croatia, and North Macedonia illustrates varying stages of development influenced by historical, economic, and policy contexts. By applying social mechanism theory, this study has provided a deeper understanding of how different contextual factors, processes, and outcomes interact to shape the success of e-health initiatives in these countries.

Slovenia and Croatia, with their more advanced and integrated digital health systems, demonstrate the importance of promoting innovation, fostering a culture of acceptance, and maintaining regulatory flexibility. Social mechanism theory has been instrumental in identifying these key drivers by highlighting the interplay between strategic planning, stakeholder engagement, and adaptive governance.

In contrast, North Macedonia, still in the earlier stages of health digitization, faces challenges that are informed by the social mechanisms at play in its unique context. The theory has helped identify the critical need for comprehensive strategic

planning, investment in infrastructure, and efforts to increase IT literacy and data standardization.

The recommendations derived from this analysis—such as promoting e-health awareness, fostering a culture of acceptance, ensuring regulatory flexibility, and learning from more developed systems—are rooted in the social mechanisms that have proven effective in Slovenia and Croatia. By adopting and adapting these mechanisms, North Macedonia can enhance its e-health system, making it more robust, efficient, and accessible.

Ultimately, the application of social mechanism theory in this comparative study not only illuminates the underlying processes driving e-health success in these countries but also provides actionable insights for North Macedonia and similar contexts. The experiences of these three nations underscore the critical importance of aligning context, processes, and outcomes through strategic planning, investment, and continuous improvement in digital health initiatives.

References

[1] Melloni, E., Pesce, F., Vasilescu, C.: Are social mechanisms usable and useful in evaluation research? Evaluation **22**(2), 209–227 (2016)

[2] The World Bank: Population in Slovenia (2023). https://data.worldbank.org/indicator/SP.POP.TOTL?locations=SI

[3] World Health Organization: Integrated, person-centred primary health care produces results: case study from slovenia. World Health Organization (2020)

[4] European Observatory on Health Systems and Policies: Health systems in transition, slovenia health system review 2021 **23** (2021)

[5] Albreht, T., Pribaković Brinovec, R., Jošar, D., Poldrugovac, M., Kostnapfel, T., Zaletel, M., Panteli, D., Maresso, A., World Health Organization: Slovenia: Health system review. Health Systems in Transition (2016)

[6] Stepovic, M., Rancic, N., Vekic, B., Dragojevic-Simic, V., Vekic, S., Ratkovic, N., Jakovljevic, M.: Gross domestic product and health expenditure growth in balkan and east european countries—three-decade horizon. Frontiers in Public Health **8**, 492 (2020)

[7] Bank, T.W.: National development strategy Croatia 2030 policy note: health sector. The World Bank, Washington, DC (2019)

[8] The World Bank: Current health expenditure ((2023). https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=HR

[9] Bank, W.: Current health expenditure (% of GDP) – North Macedonia. https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=MK&view=chart

[10] WHO-Europe: Primary health care organization, performance and quality in North Macedonia. https://www.who.int/europe/publications/m/item/primary-health-care-organization-performance-and-quality-in-north-macedonia-(2019) (2019)

[11] Vintar, M.: Development of E-Health in EU – the case of Slovenia (2019). https://www.espon.eu/sites/default/files/attachments/03%20Development%20of%20E-Health%20in%20EU%20%28Slovenia%29%20-%20Prof.dr_.%20Mirko%20Vintar.pdf

[12] Albreht, T., Polin, K., Brinovec, R.P., Kuhar, M., Poldrugovac, M., Rehberger, P.O., Rupel, V.P., Vracko, P.: Slovenia: health system review. Health Systems in Transition (2021)

[13] ZZZS: Slovenia, Timeline Milestones (2023). https://www.zzzs.si/en/slovene-health-insurance-card/timeline-milestones/

[14] Iljaž, R.J., Meglič, M., Švab, I.: Building consensus about ehealth in slovene primary health care: Delphi study. BMC Medical Informatics and Decision Making **11**(1), 1–10 (2011)

[15] Vončina, L., Arur, A., Dorčić, F., Pezelj-Duliba, D.: Universal health coverage in croatia (2018)

[16] European Observatory on Health Systems and Policies: Croatia, health system review **2** (2021)

[17] Vončina, L., Rubil, I.: Can people afford to pay for health care? new evidence on financial protection in croatia (2018)

[18] Belani, H.: Implementation of eHealth in Croatia. http://ahmevent2015.ifc.cnr.it/slides/belani.pdf

[19] Stanimirović, D., Drev, M., Rant, : Digital transformation as one of the instruments for overcoming the public health crisis: The role and use of ehealth solutions during the covid-19 pandemic in slovenia. Medicine, Law Society **15**(1), 169–192 (2022)

[20] European Commission: eHealth – Future Digital Health in the EU (2019). https://www.espon.eu/sites/default/files/attachments/Final%20report.%202019%2003%2025_final%20version_0.pdf

[21] Džakula, A., Sagan, A., Pavić, N., Lončarek, K., Sekelj-Kauzlarić, K., World Health Organization: Croatia: health system review. Health Systems in Transition (2014)

[22] Sambunjak, D., Džakula, A., Bilas, V., Erceg, M., Prena Trupec, T., Pulanić,

D., Lončarek, K.: National Health Care Strategy 2012.-2020. Government of the Republic of Croatia, Ministry of Health (2012)

[23] World Health Organization Regional Office for Europe: FROM INNOVATION TO IMPLEMENTATION: eHealth in the WHO European Region. https://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf (2016)

[24] Velinov, G., Jakimovski, B., Lesovski, D., Panova, D.I., Frtunik, D., Kon-Popovska, M.: Ehr system mojtermin: Implementation and initial data analysis. In: Studies in Health Technology and Informatics, vol. 210, pp. 872–876 (2015). https://doi.org/10.3233/978-1-61499-512-8-872

[25] Government, M.: Current Level of Implementation of e-Health Services in the Republic of Macedonia. https://proceedings.ictinnovations.org/attachment/paper/407/current-level-of-implementation-of-e-health-services-in-the-republic-of-macedonia.pdf

[26] Ampovska, M., Misheva, K.: Legal aspects of ehealth development in north macedonia. Vestnik of Saint Petersburg University. Law **3**, 660–675 (2021) https://doi.org/10.21638/spbu14.2021.311

[27] Adamič, , Eržen, I. (eds.): 30 Let Slovenskega Društva Za Medicinsko Informatiko: [publikacija Ob 30-letnici Slovenskega Društva Za Medicinsko Informatiko]. Slovensko društvo za medicinsko informatiko, ??? (2018).

[28] Rant, , Stanimirović, D., Janet, J.: Functionalities and use of the zvem patient portal and the central registry of patient data. In: 35th Bled eConference Digital Restructuring and Human (Re)action (2022)

[29] Health, C.M.: Successful completion of the project "Support for the Development of the Croatian e-Health Strategic Development Plan 2020-2025 and action plan 2020-2021". https://www.teched.hr/News/Show/36 (2022)

[30] Trupec, T.P., Maarić, M., Šajn, J., Kern, J., Pale, P., Kezdorf, V., Belani, H., Sambunjak, D., Pristaš, I., Beslač, D.D., Cega, A.C., Lazić, G., Ljubi, I., Pezo, O.: Strategic Plan of eHealth Development in the Republic of Croatia. https://www.bib.irb.hr/772627 (2014)

[31] Tomi, D.: Investigation of national readiness for e-Health in a South East European country: technology acceptance for electronic health records. https://etheses.whiterose.ac.uk/22172/ (2018)

[32] Gavrilov, G., Trajkovik, V.: New model of electronic health record: Macedonian case study. Journal of Emerging Research and Solutions in ICT **1**(2), 86–99 (2016)

[33] Government, M.: Digital Government Factsheets – Republic of North Macedonia. https://joinup.ec.europa.eu/sites/default/files/inline-files/Digital_Government_Factsheets_North_Macedonia_2019.pdf

[34] World Health Organization Regional Office for Europe: Health Systems in Action: North Macedonia. https://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf (2021)

MKQR Bill Standard and its Application on Mobile Banking and e-Invoicing

Ilija Jolevski [1][0000-0003-2262-9638], Natasha Blazeska-Tabakovska [2][0000-0002-6796-7190],
Snezana Savoska [3][0000-0002-0539-177] ,
Blagoj Ristevski[4][0000-0002-8356-1203] and Andrijana Bocevska [5][0000-0001-8701-0700],

[1,2,3,4,5] Faculty of Information and Communication Technologies – Bitola,
University "St. Kliment Ohridski" – Bitola, ul. Partizanska bb 7000 Bitola
Republic of North Macedonia

Abstract. Digital transformation is closely related to business development in a systemic way that requires knowledge and skills management and covers multiple processes including online payment. The QR code payment model for mobile banking can be widely used as an alternative to cash payment via mobile phone. An MKQR standard and application for generating MKQR codes is proposed in this paper, and it can be used as an alternative payment system and it can be integrated with the account of the source of the fund without the need to top up the transfer. The proposed system is not only meant for merchant payment but can also be used for person-to-person money transfers. This paper proposes a Macedonian open and independent standard called MKQR, using the QRcode - ISO/IEC18004 standard, for coding and data transmission methods for financial transactions.

Keywords: QR code, Mobile Banking, Mobile Payment, e-Invoicing, QR bill, MKQR standard, MKQR application

1. Introduction

Today, the most important questions are how to optimize existing business operations or functions using information technology, the cloud or digital services, and how to identify, create and manage the needs of the coming digital society. The importance of customer centricity and the identification of upcoming needs are further enhanced by the great agility and fast development of new or adjusted needs [1]. Digitalization offers huge potential to streamline business-to-business (B2B), business-to-customer (B2C), business-to-government (B2G), customer-to-business (C2B), customer-to-government (C2G), processes in terms of the manual amount of work, material, cost and time.

Digitizing an existing business process often proves to be more complicated than expected. One of the components of digitization of the business process is the transition from a paper-based process of handling invoices to electronic invoicing, as preparation for digitalization of all payments. Electronic invoicing (e-invoicing) is the exchange of an electronic invoice document, transmitted and received in a structured data format that enables automatic and electronic processing, as defined in Directive 2014/55/EU.

Due to a lack of standardization, invoices to customers were delivered in different data formats not suitable for digitalization or automation. In March 2016, the European

Committee for Standardization (CEN) approved the first steps towards a unified European standard for e-invoicing [2]. It was a grand step to payment automation and payment process streamlining.

A structured e-invoice contains supplier metadata in a machine-readable format, which can be automatically imported into the customer's billing (AP) system without requiring manual entry [3]. E-invoices contain the data only in a structured form and can be automatically imported into AP systems. With such e-bills, invoices are delivered directly to e-banking platforms, from where they can be paid in just a few clicks with full control over all transactions.

The facilitation of e-payment and e-invoicing requires connecting companies, service providers, government authorities and community researchers. Also, it should be kept in mind that legal frameworks differ worldwide, so this article is focused mostly on e-invoicing in the Republic of North Macedonia considering the e-bill Switzerland example.

Switzerland is a leader in e-invoicing and goes a step further by using a new way of invoicing for simpler payment. Switzerland has replaced traditional payment slips with mandatory QR codes [4]. The QR code included in every e-bill (printed or digital only) contains all the necessary payment information, it can be used physically on paper or digitally.

The future of m-payments is promising, considering the high rate of penetration of mobile devices, especially mobile phones, PDAs and other devices [5]. M-invoicing and QR accounts have gained significant momentum in recent years from a business perspective, as well as from governments around the world, and represent a challenging area in which digitization can be explored.

This paper proposes a Macedonian open and independent standard dubbed MKQR for coding and data transmission methods for financial transactions. This standard defines the data format that describes each aspect of an MKQR entity. Entity data is defined as plain text in MKQR format that follows the URI standard. The text is visually encoded into a standardized MKQR (quick response) code, based on the ISO/IEC18004 standard, which can then be printed on paper, displayed on a screen, etc. A code scanned by a smartphone application creates a text link containing all aspects of an MKQR entity defined by this standard.

The remainder of this paper is organized as follows. First, a review of the literature associated with m-payment and QR bills is discussed. Next, the proposed MKQR Standard is described in detail. Then the application for MKQR code generating is presented. Finally, concluding remarks are given in the last section.

2. Related Work

The four pillars of payment through a mobile application also known as Mobile Payment are defined in [6]: Self-Paying: intended for transfer to the bank account itself through mobile deposit and funds transfer capabilities feature; Paying Other People:

uses Person to Person Payment (P2P) features for individual or group payments; Paying Biller: making a payment to the biller through a mobile application owned by financial institutions or applications owned by the biller; Paying Merchant/Retailer: is for payment transactions on purchases at merchants using NFC sensors, QR code, cloud, or online.

Based on the mapping of the four pillars of mobile payment strategy from Sachdev, the goal of QR Code Payment in mobile banking is to build an ecosystem that will facilitate payment/issuing and ultimately reduce the volume of cash withdrawal transactions. It can be used as an alternative method of payment transactions with a merchant and for service person to person. Building an ecosystem is a long-term process.

QR Code Payment is based on QR (quick response) codes, that represent 2D matrix-type symbols with a cell structure arranged in a square. QR codes were approved as an ISO international standard (ISO/IEC18004) in June 2000. Unlike many traditional 2D bar codes that need to be decoded by a specific scanner, QR codes can be decoded by a small program in a cell phone or a personal computer with a built-in camera. According to the security of the online transaction, QR code as well as NFC and biometrics is categorized as an alternative to the multi-factor authentication process. In addition, QR code only needs a camera to scan, so all types of mobile phones have QR Code support. Although initially, QR codes were mainly used in areas such as downloading digital content and product information, today, they are applied in various application streams related to marketing, security, academics, as well as finance [7].

EU has set the goal to make e-invoicing the primary method of invoicing by 2020 [8]. Therefore, the EU launched initiatives to drive the adoption of e-invoicing such as the Pan-European Public Procurement OnLine (PEPPOL) initiative seeking to enable interoperability between dissimilar systems by providing technical specifications being implemented into existing e-procurement applications [9]. Moreover, the United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT), a body of the United Nations Economic Commission for Europe (UNECE), drives the harmonization of trade processes internationally [2].

Today, from an EU legal point of view, sending documents in a PDF format via e-mail is a legally compliant way of e-invoicing, as long as the authenticity and integrity of the document can be proven through adequate internal business controls and it is properly supported by accounting documents.

From a technological perspective, there has been a recent convergence of e-invoicing standards, across different solutions and systems. Moreover, there is a trend towards business software solutions implementing e-invoicing as a basic functionality, following established standards [2].

Switzerland is going a step forward and a new way of invoicing use for simpler of payment. This new option is based on the standards of the Single Euro Payments Area (SEPA), which harmonize bank transfers across 34 states in Europe and was established using a unified data format following the ISO 20022 standard [10]. In 2017, the Swiss Financial Center presented the new QR bill, a future-oriented solution that enables the different interest groups to meet the challenges of digitalization and regulation

efficiently. The QR bill was introduced in 2020 as part of efforts to harmonize and digitalize the Swiss payment transactions ecosystem. After a two-year transition period, all financial institutions in Switzerland will discontinue processing red and orange payment slips. The QR bill replaces traditional payment slips [4]. QR code contain all necessary payment information, it can be used physically on paper or digitally.

QR-code, also displays the payment information in text format on the right, making it readable for non-automated processing as well. The invoice-processing software will inevitably have to incorporate relevant functionalities for generating invoices and reading the new payment slip QR bill.

QR bill or digitization of invoicing allows faster payment and fewer errors without typing the account numbers and reference numbers. The QR bill, defined according to ISO 20022 standards, facilitates payments thanks to automated data processing. This is not just SEPA-conforming, in terms of euro payments, but accommodates all domestic and foreign payment transactions [7]. It enables both issuing and paying invoices.

Since 2017, Switzerland used the new QR bill, which can be printed or issued digitally and is a replacement for previous payment slips. The QR bill consists of a payment and confirmation section. All payment information is contained both digitally in the Swiss QR code and - as usual - in plain text. This allows the recipient of the invoice to check the accuracy of the payment data after scanning and before approving the payment and, if necessary, enter the payments manually. The QR bill is divided into two parts (like the existing payment slips): 1) a confirmation part and 2) a payment part. The Swiss QR code contains all the relevant information needed for invoicing and payment. The QR bill is perforated, so the recipient of the invoice can easily separate the payment and bill part of the invoice.

Modern m-payment systems, from the traditional “buyer-seller” exchange, passed to more complex transaction models, including also network providers, finance companies for money transaction management, generally a credit card or debit card issuing company and/or other institutions operating inside the Internet [11]. Taking into account the right end users, the right payment circuit, technology factors and business models can produce a successful m-payment solution. Security requirements must be imposed, in order to ensure trust and avoid fraud, as well as interoperability and privacy requirements. In addition, in this m-payment environment, the speed of execution and ease of use are mandatory requirements, too.

3. MKQR Standard

In terms of PESTEL (political, economic, social, technological, environmental and legal) factors, it becomes obvious that the political, technological and legal infrastructure in the Republic of North Macedonia is ready to support digitalization and that a great effort has been made to facilitate digital changes. Although not all factors are truly optimal, from an economic perspective, the business case for digitizing business processes is strong. Since there will never be an ideal moment in which there is absolute certainty concerning all circumstances, the digitization of business processes

including the digitization of all types of financial obligations, should be understood as a basic task of managers in the Republic of North Macedonia.

The MKQR is an independent and open proposed standard, based on the ISO/IEC18004 standard [12], for coding and data transmission methods for financial transactions. By simply scanning the QR code with his mobile phone in the payment section and approving the payment, without the need for additional entries in e-banking, the recipient of the bill will be able to complete the entire transaction. The standard defines the form of data that describes every aspect of an MKQR entity. The data for an entity is defined as plain text in MKQR format that follows the URI standard. The text is visually encoded into a standardized MKQR, which can then be printed on paper, displayed on a screen, etc. A code scanned by a smartphone application creates a text link containing all aspects of an MKQR entity defined by this standard. Applications that fully implement this standard can perform financial transactions through standardized text links and their corresponding MKQR codes. The standard is versioned semantically. The base version is 1.0.0. The definitions of data, the description of process and the plain text format are given in Table 1.

Table 1. Definitions of data attributes describing a payment process and the plain text format

Term	Description
MKQR	Reference to this standard
MKQR URI	Plain text following the URI standard
Entity	An entity consists of all the data defined by the MKQR standard
Client	An application that implements the MKQR standard
Creditor	The private or legal entity that is the entity's destination
Debtor	The private or legal entity that is the source of the entity

The data for each entity is combined into a single plain text instance in *MKQR URI format* that applications implementing the standard can parse. The MKQR URI format starts with a fixed segment: mkqr://pay?. All required data and all data containing the MKQR URI text must be valid. Validation is done before generating the MKQR code. MKQR does not guarantee the complete validity of the data as it cannot be fully validated before the generation of the MKQR code. Customers can perform additional data validation, according to their own needs.

The MKQR URI starts with the “mkqr://” prefix to be possible to be used as a trigger to activate and start any smartphone app that can handle such URIs. The idea is that the standard does not define the apps, there can be multiple apps that support and recognize MKQR and it will be up to the user to choose which app will be used to handle the MKQR data. In a practical scenario, most probably the first adopters of such support will be the e-banking apps of the banking providers (banks).

The MKQR code can be in color or monochrome. A color QR code has a gradient that starts at the top of the red QR code (#CC0708) and ends at the bottom of the black QR code (#000000). In the center is the MKQR code logo. The MKQR code logo has a 1-square-thick border on all 4 sides. The frame has a yellow color defined on the flag of the Republic of North Macedonia. In the center of the logo are the letters MK, colored with the yellow color defined on the flag of the Republic of North Macedonia,

which has a total width of 11 squares and a total height of 5 squares. The size of the logo is 13 squares in width and height. The background is red as defined on the Macedonian flag. The logo transparency is 20%.

A monochrome code contains only the colors black and white. The dimensions of the logo are the same as the dimensions of the color QR code. The background of the monochrome code is black, and the frame and text are white. Examples of the appearance of the MKQR code are illustrated in Figure 1.

Figure 1: Examples of the appearance of the MKQR code.

Table 2 contains explanations of all elements that are defined by the MKQR standard. Table 2 shows all entity data of the attributes included in the MKQR string (abbreviations, definitions, values, data types and some remarks), with the following abbreviations in the data type column: M - Mandatory; O – Optional; C – Conditional; S - structured address; K - combined address.

Table 2. Entity data

Attribute Name	Abbr.	Definition	Values	Note	Data Type
QRType	t	Standard type.	MKD	MKD is a fixed value for MKQR.	M
Version	v	Version of the specification according to which the MKQR code is generated.	Fixed length, 4 digits.	1.0.0	M
Coding	c	Type of code page of the data in the text.	1 for UTF-8 encoding but only with a subset of Latin characters. 2 for UTF-8 encoding with all characters including Cyrillic.		M
IBAN	iban	IBAN of the creditor's account.	Complex data is validated with the expression.	Customers can also implement	M

Attribute Name	Abbr.	Definition	Values	Note	Data Type
				additional IBAN validations.	
Alternative IBAN	aiban	Alternative IBAN accounts of the creditor, separated by pipe character "|".	Maximum 77 characters, each individual entry is validated with the expression.	Alternative accounts are arranged according to the creditor's preference.	O
Address Type	cat	Creditor Address Type.	S - structured address in multiple fields. K - combined address in two fields.		M
Creditor Name	cn	Name (and surname) of the creditor.	Maximum 70 characters.		M
Street or address line 1	cadd1	For a structured address S, the creditor's street name or postal code is entered. For combined address K, the first line of the address, street and number or postal code is entered.	S - maximum 16 characters. K - maximum 70 characters.	If cat=K, then this data must exist.	O
Building number or address line 2	cadd2	For a structured address S, the address number of the creditor is entered. For combined address K, the second line of the address is entered: the creditor's city and postal code.	S - maximum 16 characters. K - maximum 70 characters.		O
Postal Code	cz	For a structured address S, the postal code is entered.	S - maximum 7 characters. K - is omitted.	If cat=S (structured address), then this data is required.	C
Town Name	cg	The name of the creditor's city.	S - maximum 35 characters. K - is omitted.	If cat=S (structured address), then this data is required.	C
Country	cc	State of creditor.	Fixed length, 2 characters, according to the ISO 3166-1 standard.		M
Amount	a	Amount	An 8-byte decimal numeric data equivalent to a double as defined in the IEEE-754 standard.	The sum must be a text that can be converted to a double. If the amount cannot be converted into this format the MKQR generation should be stopped.	O

Attribute Name	Abbr.	Definition	Values	Note	Data Type
Currency	cur	Payment currency.	Fixed length, 3 characters, according to the ISO 4217 standard.	The customer determines whether the currency of payment is allowed. Whether the currency is valid is determined by its presence.	M
Ultimate Debtor address type	pat	The debtor's address type.	S - in multiple fields. K - in two fields.		O
Ultimate Debtor Name	pn	Name (and surname) of the debtor.	Maximum 70 characters.		O
Street Name or Address Line 1	padd1	For S, the creditor's street name or postal code is entered. For K, the first line of the address, street and number or postal code is entered.	S - maximum 16 characters. K - maximum 70 characters.		O
Building Number or Address Line 2	padd2	For a structured address S, the address number of the creditor is entered. For combined address K, the second line of the address is entered: the creditor's city and postal code.	S - maximum 16 characters. K - maximum 70 characters.		O
Postal Code	pz	For a structured address S, the postal code is entered.	S - maximum 7 characters. K - is omitted.	If pat=S (structured address), then this data is mandatory.	C
Town Name	pn	The name of the creditor's city.	S - maximum 35 characters. K - is omitted.	If pat=S (structured address), then this data is mandatory.	C
Country	pc	State of creditor.	Fixed length, 2 characters, according to the ISO 3166-1 standard.	Whether a country code is valid is determined by attendance	O
Payment Reference Type	rt	Reference type. The following codes are allowed: QRR – QRR reference. SCOR – Creditor Reference (ISO 11649) NON – no reference.	Maximum of 4 alphanumeric characters. QRR (when using QR-IBAN) or SCOR / NON (when using IBAN).		M
Payment Reference	ref	Reference. A structured reference is either a QRR reference or an ISO 11649 - Creditor reference.	Up to 27 alphanumeric characters; QRR reference: 27 numeric characters, calculation of calculation sum according	Mandatory if QR-IBAN is used.	C

Attribute Name	Abbr.	Definition	Values	Note	Data Type
			to module 10 recursive, 27th position of the reference. Creditor reference (ISO 11649): maximum 25 characters, alphanumeric. Ignored if filled for rt=NON.		
Payment Code	pcd	Payment code	3 digits	The description of the codes is available.	M
Payment Type	nac	Method of payment	1 digit		O
PP50 Payment account	us50	Consignee's transaction account	Fixed length, 15 digits.		O
PP50 Account Single User	usek50	Consignee's transaction account	Fixed length, 15 digits.		O
PP30 Payment account	us30	Consignee's transaction account	Fixed length, 15 digits.		O
PP30 Account Single User	usek30	Consignee's transaction account	Fixed length, 15 digits.		O
Additional info, USTRD	i	Additional information	A maximum of 140 alphanumeric characters.	Documentation	O
CheckURL	curl	Address for additional validation of entered data. The response must be HTTPS in JSON format and must contain an "IsValid" attribute that defines whether the validation was successful or not.	Complex data is validated with the expression		O
Alternative Payment Scheme	ap	Name of the alternative payment method.	Maximum of 20 alphanumeric characters.		O
Alternative Payment Values	av	Value of the alternative payment method.	An 8-byte decimal numeric data equivalent to a double as defined in the IEEE-754 standard.		O
Alternative Payment Description	ad	Description of the alternative payment method.	Maximum of 240 alphanumeric characters.		O

Attribute Name	Abbr.	Definition	Values	Note	Data Type
Alternative Payment Currency	ac	Currency of the alternative payment method. It can be omitted.	Fixed length, 3 characters, according to the ISO 4217 standard.	Whether the currency of payment is allowed is determined by the customer. Whether the currency is valid is determined by its presence.	O

An example of MKQR encoded string for e-invoice payment with multiple destination IBAN accounts and metadata:

mkqr://pay?t=1&v=200&c=1&iban=3123452523424&aiban=210068329110129|3000000003538887|530010101565335|200002545849091|250010000232377|270068329110136&cat=1&cn=Топлификација Скопје&cadd1=Лондонска бр. 8&cadd2=&cz=1000&cg=Скопје&cc=Македонија&a=1751,00&cur=MKD&pat=&pn=Петко Петков&paddr1=Партизанска 111&paddr2=&pz=1000&pg=Скопје&pc=Македонија&rt=&ref=17180904216&i=&=&=&ap=&av=&ad=&ac=&pc=&nac=&us=&usek=&ps=&pr=

4. Creation of MKQR strings

The application together with the invoice form is designed to work collaboratively with other systems to which it will be connected. The application is designed to be filled out automatically by the system to which it will be connected. The QR code is generated and the user is redirected to the website where he will pay the bill.

After the invoice is paid, the invoice application provides the data about the user and the amount he paid, together with the same QR code that he scanned in PDF format, with the possibility to download and save it, for further records. In this way, the user can whenever he wants to scan the QR code and find out when, where, how much and what he paid for that amount of money.

The application consists of a web form that has 37 fields to fill. Each place is specific to a certain data. The application detects whether numbers or letters are written in the fields, and whether the language entered is English or Macedonian, Cyrillic or Latin. In the process of code generation, if an important field is empty, it warns the user with a message that the field is mandatory and needs to be filled (Figure 2). When the necessary and mandatory fields are filled in, the QR code appears. The code is intended for charging for services and products in Macedonia, where through it, by scanning it, citizens will be able to pay bills, donations, products and services from home and via smartphone or computer. The QR code is specific to Macedonia with a logo with the initials MK or the Macedonian flag in the middle of the QR code (Figure 1). When the

QR code is scanned, the user is redirected to the specified website to complete the transaction. After that procedure is completed, the Invoice application appears which provides the invoice in PDF format with the data entered in the form.

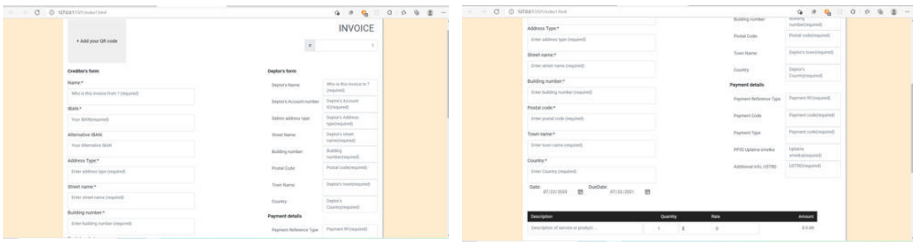


Figure 2. MKQR Application – QR code generation

For the application development, several technologies and tools were used: JavaScript, NodeJS, HTML2PDF and jsPDF - an open-source library for generating PDF documents using JavaScript.

5. Concluding Remarks

The MKQR standard and prototype of the MKQR application will modernize payment transactions. MKQR invoices will offer many advantages to both recipients and issuers of invoices. They are convenient because scanning the QR code is so simple. The new account will be both fast and efficient. A few clicks are all it takes to release payment, replacing the tedious task of typing in invoices and reference numbers and significantly reducing a common source of error.

At the same time, the proposed MKQR standard and prototype of the MKQR Application bridges the digital and paper-based worlds and goes a step further. It will advance the digitization processes of businesses in the North Republic of Macedonia. The proposed MKQR standard and MKQR Application will offer many advantages for invoice recipients and invoice issuers such as all payment information transferred electronically; all payment information built into a QR code in digital form; one QR code for all types of payment and reference; more straightforward invoice processing; simplified payment reconciliation, less manual work, fewer errors etc.

The new m-payment infrastructure will ensure a seamless digital payment experience.

References

[1] D. Schwaferts and S. Baldi, "Digital Transformation Management and Digital Business Development," in *Business Information Systems and Technology 4.0: New Trends in the Age of Digital Change*, Basel, Switzerland, Springer, 2018, pp. 81-102.

[2] C. Tanner and S.-L. Richter, "Digitalizing B2B Business Processes—The Learnings from E-Invoicing," in *Business Information Systems and Technology 4.0: New Trends in the Age of Digital Change*, Basel, Switzerland, Springer, 2018, pp. 103-116.

[3] European Commission, "What is eInvoicing," 06 2024. [Online]. Available: https://ec.europa.eu/digital-building-blocks/sites/display/DIGITAL/What+is+eInvoicing.

[4] SIX Interbank Clearing Ltd 2017, "The End of Swiss Payment Slips: How to Keep Your Standing Orders Activated and Working," 06 2024. [Online]. Available: https://www.six-group.com/en/blog/qr-bill-standing-orders.html.

[5] L. Antovski and M. Gusev, "M-Payments," in *2nd Int. Conf. lnformafion Technology lnferfaces /TI 2003*, Cavtat, Croatia, 2003.

[6] S. Sachdev, "(2014). The Four Pillars of Mobile Payments - Immediate Opportunities.," 2014.

[7] S. Tiwari, "An Introduction To QR Code Technology," in *2016 International Conference on Information Technology, (ICIT),* Bhubaneswar, India, 2018.

[8] European Commission 2013, "E-invoicing in public procurement: another step towards end-to-end e-procurement and e-government in Europe," 06 2024. [Online]. Available: http://europa.eu/rapid/press-release_IP-13–608_en. html.

[9] Pan-European Public Procurement OnLine (2017), "What is PEPPOL? " 06 2024. [Online]. Available: http://peppol.eu/what-is- peppol.

[10] The National Adherence Support Organisation (NASO) of Switzerland 2017, "Single Euro Payments Area (SEPA)," 06 2024. [Online]. Available: https://www.sepa.ch/en/home/sepa.html.

[11] A. Vizzarri and F. Vatalaro, "m-Payment systems: Technologies and business models," in *2014 Euro Med Telco Conference (EMTC)*, Naples, Italy, 2014.

[12] "ISO/IEC 18004," 07 2024. [Online]. Available: https://www.swisseduc.ch/informatik/theoretische_informatik/qr_codes/docs/q.

[13] "Six," 07 2024. [Online]. Available: https://www.six-group.com/en/products-services/banking-services/payment-standardization.html.

[14] T. J. Soon, "QR Code," *Synthesis journal,* pp. 59-78, 2008.

[15] T. S. Parikh and E. D. Lazowska, "Designing an architecture for delivering mobile information services to the rural developing world.," in *Designing an architecture for delivering mobile information services to the rural developing world.*, 2006.

[16] Seino et al., "Development of the traceability system which secures the safety of fishery products using the QR Code and a digital signature," in *Marine Technology Society/IEEE TECHNO-OCEAN*, 2004.

[17] Ruslan, G. M. Karmawan, Suharjito, Y. Fernandoand and A. Gui, "QR Code Payment in Indonesia and Its Application on Mobile Banking," in *KnE Social Sciences / FGIC 2nd Conference on Governance and Integrity*, 2019.

[18] M. V. Stiphout and M. Mual, "Payment Methods Report 2017.," in *The Paypers BV.*, 2017.

Session 7

Beyond Single-Label Classification: Enhancing Radiological Diagnostics for Thoracic Diseases with Azure AutoML

Olga Petan[1], Aleksandar Karadimce[1][0000-0002-5013-7967], Dijana Capeska Bogatinoska[1][0000-0001-5474-8290] and Ljubinka Sandjakoska[1]

[1] University for Information Science and Technology "St. Paul the Apostle"
Ohrid, Republic of North Macedonia
olga.petan@isvma.uist.edu.mk, aleksandar.karadimce@uist.edu.mk,
dijana.c.bogatinoska@uist.edu.mk, ljubinka.gjergjeska@uist.edu.mk

Abstract. The global COVID-19 pandemic has catalyzed significant expansion in artificial intelligence (AI) research within healthcare, highlighting the challenges humanity faces in managing large-scale health crises. In response, research teams worldwide have employed computer vision to predict COVID-19 presence and severity. Despite these efforts, the clinical utility of AI models remains limited, as they do not adequately address common errors made by radiologists: satisfaction of search, premature closure, and anchoring bias. Most AI research has focused on single-label classifiers and severity predictors, which we identify as problematic, as they cannot prevent missed diagnoses or identify additional diseases. Radiologists can detect moderate and severe thoracic diseases, but certain conditions may not be visible on X-rays in patients with mild symptoms. Hence, the added value of these AI models in clinical settings is unclear. Moreover, accuracy metrics can be misleading without radiologist-annotated bounding boxes outlining pathological areas. We address this by employing the intersection over union (IoU) metric, a spatial metric that indicates the portion of the observed pathological area identified by the model. This research utilizes X-ray images from patients with healthy lungs or various lung diseases, and Microsoft Azure AutoML to develop a multi-output classification model. Our model mitigates known cognitive errors in radiology by indicating the number of diseases, naming them, and highlighting pathology locations. Finally, we propose a comprehensive diagnostic system incorporating medical imaging as one component. To create clinic-ready AI products, we emphasize the necessity of diverse, well-labeled data spanning various demographics.

Keywords: multi-label classification, Azure AutoML, computer vision, AI in healthcare, healthcare diagnostics.

1 Introduction

The field of artificial intelligence (AI) experienced a significant surge during the COVID-19 pandemic, with computer vision and neural networks being extensively utilized to develop models aimed at easing the burden on the medical sector and aiding in triage and treatment. Predominantly, these use cases involved single-output classifiers that predict one disease and classifiers that assess the severity of a patient's condition. However, AI models that predict only one disease have limited clinical utility for several reasons: patients often suffer from multiple diseases simultaneously, necessitating a model that can identify all present conditions. For instance, the morbidity of COVID-19 is often influenced by comorbidities such as Parkinson's disease, diabetes, cancer, and other lung diseases. Additionally, while human radiologists can distinguish between moderate and severe lung diseases in medical images [1], they are prone to cognitive errors during the diagnostic process. Given the limited availability of antiviral treatments that effectively differentiate between viruses, even a highly accurate model distinguishing viral lung diseases, such as coronaviruses and influenza, may offer limited clinical benefit if treatment protocols remain the same.

Considering the psychological nature of radiologists' diagnostic errors, an AI model that addresses these mistakes could provide a valuable solution. The three most common cognitive errors made by radiologists are satisfaction of search, premature closure, and anchoring bias. Radiologists typically review a patient's health records before examining medical images, which can lead to cognitive biases. Satisfaction of search occurs when radiologists look for diseases that match the health record information, potentially overlooking additional abnormalities. Premature closure happens when radiologists, often pressed for time, settle on the most obvious abnormality, influenced by prior information. Anchoring bias arises when radiologists fail to integrate new or contrasting information into their diagnostic process.

This paper presents a multi-label lung disease classifier designed to assist radiologists by offering flexible outputs. The classifier can indicate the presence of pathology in X-ray images, enumerate the detected diseases, identify the specific diseases, and highlight the locations of the pathologies on the images.

2 Related work

2.1 Single-label lung disease classification

Researchers have employed various techniques to develop single-label lung disease classifiers. An ensemble of deep convolutional networks utilizing MobileNet, DenseNet, and Vision Transformer [2] achieved an accuracy of 93.91% and an F1-score of 93.43%. In [3], researchers used both normal and segmented images trained on four networks—SqueezeNet, ResNet18, Inceptionv3, and DenseNet121—to differentiate between COVID-19, MERS, and SARS. The most effective network produced a model with an accuracy of 98% and an F1-score of 97.7%. In [4], authors developed Detail-Oriented Capsule Networks (DECAPS), a network that automatically diagnoses

COVID-19 computed tomography (CT) scans. The study first segments the region of interest using inverted dynamic routing, and then generates activation maps using a two-stage patch crop and drop strategy. The network was able to achieve 84.3% precision, 91.5% recall, and a 96.1% area under the curve [4]. The authors in [5] developed an attention-guided convolutional neural network consisting of three branches: the first branch learns the regions specific to a disease and discards the noisy regions, the second branch crops these regions, and the third branch is used to train the ResNet50 model. Authors report an area under the curve (AUC) of 86.8%[5].

2.2 Multi-label lung disease classification

In [6], researchers developed a hybrid network architecture consisting of a spatial encoder module, a context encoder module, and a multi-branch output layer to train a multi-label classifier on the same dataset used in this paper. They compared their model to top-performing models such as ConsultNet, LLAGnet, and TNELF, demonstrating that their hybrid network consistently outperformed these models, achieving an average AUC of 83.8% across all labels [6]. The impact of class imbalance, where certain diseases lack sufficient images, on the model's performance is highlighted in [7]. Researchers trained a multi-label classifier using five networks: CustomNet, DenseNet121, ResNet50, Inception, and VGG16. Despite achieving a relatively high reported accuracy of approximately 87%, the models struggled to predict positive cases for certain diseases, such as consolidation and pleural effusion. The authors in [8] fine-tuned the ConvNeXt network to get visual vectors, which were then combined with semantic vectors encoded by BioBert to form a single metric space. The model trained on this mapped vector achieves an AUC of 82.6% [8].

2.3 COVID-19 severity classification

Given the critical importance of assessing a patient's risk of deterioration or death during the pandemic, researchers prioritized predicting the severity of COVID-19 infections. In [8], researchers augmented the image dataset with patient information such as age, sex, and a pre-computed lung infection ratio, achieving an AUC of 79% for severe cases. Additionally, deep learning and AI were employed to develop forecasting or time series severity classifiers, which do not utilize images. In [9], researchers applied XGBoost, a Bayesian classifier, a neural network, and a support vector machine to analyze vital signs and blood analysis data from patients to predict whether a patient would become critical during their hospital stay. They reported that the neural network achieved an accuracy of 77%. In [10], authors classify CT scans into Non-COVID, Severe-COVID, and Non-Severe COVID. The images are preprocessed using Contrast Limited Histogram Equalization, and then used to train a VGG-16 network. The accuracy metrics are calculated over 10 cross-validation iterations using a support vector machine. The authors report an accuracy of 96.05% and an F1-score of 95.96%[10].

3 Dataset Construction

The images used in our research are sourced from the CXR8 NIH repository [11, 12], which comprises 112,120 images labeled with 14 disease categories and one category for healthy images. This repository includes a metadata CSV file, providing additional information about the patients' ages and sex. The datasets include the following labels: Infiltration, Atelectasis, Effusion, Nodule, Pneumothorax, Mass, Consolidation, Pleural Thickening, Cardiomegaly, Emphysema, Fibrosis, Edema, Pneumonia, and Hernia.

To reduce computational costs and minimize the time required to upload images to Azure and execute the multi-label model, we utilized approximately 25% of the available images. This approach ensures a balanced distribution of metadata and labels, providing a comparable number of healthy and pathological images, a similar distribution of view types: anterior to posterior (AP) and posterior to anterior (PA), and an equivalent number of male and female patients. The final dataset consists of 26,152 images. Table 1 presents the categorization of images by the number of diseases, with 0 indicating a healthy image.

Table 1. Number of Images per Number of Diseases Group

Number of Diseases	Number of images
0	10000
1	3836
2	6480
3	4205
4	1247
5	301
6	67
7	16
Grand Total	26152

Figure 1 shows sample images of healthy patients, patients with one disease, and patients with multiple diseases:

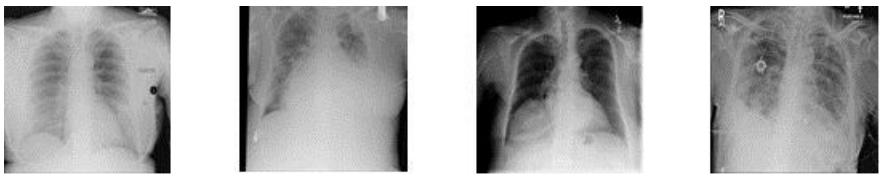


Fig. 1. (a) Healthy Image, (b) Cardiomegaly, Consolidation, (c) Atelectasis, Cardiomegaly, Fibrosis, (d) Consolidation, Effusion, Infiltration, Mass.

Table 2 displays the most frequent label combinations for each number of diseases. For instance, among images with two diseases, the most common pair is effusion and infiltration, followed by atelectasis and infiltration, and so on.

Table 2. Top 5 Most Frequent Labels per Number of Diseases Group

One disease	Two Disease	Three Diseases	Four Disease
Infiltration	Effusion, Infiltration	Atelectasis, Effusion, Infiltration	Atelectasis, Consolidation, Effusion, Infiltration
Atelectasis	Atelectasis, Infiltration	Atelectasis, Consolidation, Effusion	Atelectasis, Cardiomegaly, Effusion, Infiltration
Effusion	Atelectasis, Effusion	Edema, Infiltration, Pneumonia	Atelectasis, Effusion, Infiltration, Pneumothorax
Nodule	Infiltration, Nodule	Consolidation, Effusion, Infiltration	Atelectasis, Effusion, Infiltration, Pleural Thickening
Pneumothorax	Cardiomegaly, Effusion	Edema, Effusion, Infiltration	Atelectasis, Effusion, Infiltration, Mass

The image dataset includes two views: AP (Anterior-Posterior) and PA (Posterior-Anterior). In the PA view, the X-ray beam passes from the patient's back to the front, while in the AP view, it passes from the front to the back. There are 12,809 images captured in the AP view and 13,343 in the PA view. Table 3 provides the categorization of images categorized by view type and number of diseases detected.

Table 3. Number of Images per View Type and Number of Diseases Group

Number of Diseases	AP	PA
0	5000	5000
1	1887	1949
2	2946	3534
3	2140	2065
4	661	586
5	142	159
6	26	41
7	7	9
Grand Total	12809	13343

We obtained images from a cohort of 10,805 patients, comprising of 4,783 females and 6,022 males. Table 4 presents the categorization of male and female patients across different age ranges, while Table 5 illustrates the number of male and female patients, along with their average age categorized by the number of diseases detected.

Table 4. Number of Patients per Age Group and Sex

Age Range	F	M
< 10	69	87
10-19	218	295
20-29	488	620
30-39	684	771
40-49	1024	1068
50-59	1167	1470
60-69	789	1147
70-79	289	474
80+	55	90

Table 5. Number of Patients and Average Age per Sex and Number of Diseases Group

# Diseases	F	M
1	3704	4708
	avg.age 46.3	avg.age 47.1
2	1542	2028
	avg.age 49.3	avg.age 49
3	979	1305
	avg.age 49.4	avg.age 48.4
4	368	492
	avg.age 49.4	avg.age 48.1
5	95	146
	avg.age 49.1	avg.age 49.5
6	25	35
	avg.age 55.3	avg.age 45.8
7	3	13
	avg.age 57.3	avg.age 42

It is observed that the average age generally increases with the number of diseases detected in the image for both sexes. However, despite having only 13 images depicting 7 labels from male patients, their average age is notably lower at 42 years.

4 Experimental Methods and Results

4.1 Transfer Learning and Azure AutoML Hyperparameter Tuning

Neural networks are data-intensive and require substantial datasets for effective training. However, researchers frequently lack access to sufficiently large datasets comprising only their images. To overcome this limitation, transfer learning is employed, where a pre-trained model is fine-tuned for the specific task at hand. In this study, we utilize MobileNetv2 and ResNet152 for this purpose.

Furthermore, neural networks feature numerous adjustable parameters that influence model performance during training. Typically, these parameters are fine-tuned manually or through hyperparameter tuning techniques. Hyperparameter tuning involves methods like grid or random search, which systematically explore a range of values for each parameter and evaluate the model's performance accordingly. Table 6 presents the specific parameter values utilized across ten training iterations conducted on Azure.

We provide a concise overview of the training parameters utilized in our research work:

- *Learning rate*: This parameter governs the adjustment of model weights between successive training steps. Optimizing the learning rate is crucial for effective training; if too high, training may terminate prematurely with suboptimal solutions, whereas if too low, convergence to a global minimum may be hindered by extended training times.

- *Batch size*: Batch size refers to the number of images processed concurrently during model training. Model weights are updated after processing each batch. Common batch sizes include 32, 64, and 128. Larger batch sizes consume more memory, which can be cost-prohibitive, while smaller batch sizes may introduce noise into weight updates.
- *Epochs*: An epoch represents a single iteration of the training process over the entire dataset. During a single epoch, model weights are updated after processing each batch of images. Multiple epochs involve repeating this process iteratively. For instance, with 10 epochs, the model iteratively updates weights after each batch until all training images are processed, repeating the cycle 10 times. It is critical to carefully select the number of epochs to avoid overfitting, where excessive epochs may lead to the model fitting too closely to the training data but fail to predict correctly on new, unseen images.
- *Training and validation splits*: During training, the model's parameters are optimized based on the data it encounters. However, this may not always translate to optimal performance on unseen data, a phenomenon known as generalization. To mitigate this, datasets are typically partitioned into training and validation sets. For example, an 80% allocation to training data and 20% to validation data ensures the model's performance can be assessed on unseen data, facilitating better estimation of its generalizability.

Neural networks involve numerous parameters that must be carefully tuned to optimize model performance and ensure accurate predictions. Given the interdependencies among these parameters, exploring various combinations is essential. Manual parameter testing is typically slow and costly; therefore, we employ hyperparameter tuning techniques such as grid search and random search.

- *Grid search*: We define a grid of hyperparameter values and systematically train models on each combination. For instance, if the learning rate choices are 0.001 and 0.01, and batch sizes are 32 and 64, we train models with configurations like 0.001 learning rate and batch size 32, and 0.01 learning rate and batch size 64.
- *Random search*: This involves selecting values randomly from predefined distributions. Using the same example, a random search might explore a learning rate between 0.001 and 0.01 and batch sizes between 32 and 64, choosing combinations randomly within these ranges.

Azure AutoML is employed to automate the parameter optimization process, facilitating the discovery of optimal parameters and model configurations. The models we have available in the AutoML class for multilabel classifiers are:

- MobileNet: a model used for mobile applications
- ResNet: models using residual networks
- ResNeSt: models using split attention networks
- SE-ResNeXt50: – models using squeeze and excitation networks
- ViT: models using vision transformers networks

Gien the constraints imposed by limited computational resources, a comprehensive evaluation of all potential model configurations and hyperparameter combinations is

computationally infeasible. Further research and increased funding are warranted to facilitate a more exhaustive exploration of available model architectures.

The dataset is partitioned into a training set comprising 80% of the images and a validation set with the remaining 20%. Azure AutoML conducts training for up to 2 hours or a maximum of 15 training jobs, whichever is completed first. The image model training class within Azure AutoML incorporates standard computer vision techniques for image augmentation, including pixel value normalization, resizing, cropping, horizontal flipping, and adjustments to saturation, brightness, contrast, and hue.

Table 6 presents the parameters and their corresponding values used in each training job. The last three training jobs were terminated after approximately 30 seconds of training due to insufficient improvement in accuracy or other metrics with the chosen parameter configurations.

Table 6. Parameter Name and Parameter Value per Training Job 1 through 10

Parameter	Job 1	Job 2	Job 3	Job 4	Job 5	Job 6	Job 7	Job 8	Job 9	Job 10
Learning Rate	0.0953	0.0984	0.0186	0.0907	0.0747	0.056	0.069	0.0985	0.037	0.0232
Model name	Mobilenetv2	Mobilenetv2	Resnet152	Mobilenetv2	Mobilenetv2	Resnet152	Mobilenetv2	Resnet152	Mobilenetv2	Mobilenetv2
Optimizer	Sgd	Sgd	Adam	Adam	Sgd	Adam	Adam	Sgd	Sgd	Adam
Training crop size	256	224	224	224	224	224	224	224	256	256
Training batch size	64	64	32	64	32	32	32	32	64	64
Validation crop size	224	256	256	256	256	224	256	256	224	256
Validation batch size	288	288	320	320	320	352	288	288	320	288
Grad accumulation step	1	1	1	1	1	1	1	1	1	1
Number of epochs	15	15	6	15	15	4	15	4	15	15

4.2 Performance Measures

We are considering four distinct accuracy metrics: accuracy, AUC-weighted, weighted F1-score, and intersection over union.

- *Accuracy* measures the percentage of correctly predicted labels, defined as the ratio of correctly predicted labels to the total number of predictions. However, this metric can be less informative in multilabel scenarios, especially when the labels are imbalanced. In our model, accuracy has two notable shortcomings:
  - If an image contains multiple diseases, the model must predict the exact sequence of labels in the correct order to be considered accurate.
  - The dataset is imbalanced, with fewer than 500 images for patients with five or more diseases.

$$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions} \quad (1)$$

- *AUC (Area Under the Curve)* measures the area under the receiver operating characteristic (ROC) curve, which plots true positive rates against false positive rates at various classification thresholds. A higher AUC value, closer to 1, indicates better model performance in distinguishing labels. The weighted AUC assigns weights to the AUC values of each class based on their prevalence.
- *F1-Score* accounts for both false positives and false negatives by incorporating both precision and recall. This metric is particularly useful for uneven class distributions. We use the weighted F1-score to adjust for class imbalance, ensuring each class contributes proportionally to the overall score.

$$F1Score = 2x\frac{PrecisionxRecall}{Precision+Recall} \quad (2)$$

  - Precision is the ratio of true positives to the sum of true and false positives. It is crucial in healthcare to avoid both missing unhealthy patients and subjecting healthy patients to unnecessary tests.
  - Recall measures the proportion of actual positive cases correctly identified by the model.
- *IoU (Intersection over Union)* relaxes the strictness of the accuracy metric by evaluating the spatial overlap between the predicted and true label areas. It is defined as the ratio of the intersection to the union of the predicted and true label areas.

$$IoU = \frac{ofPredicted\wedge TrueLabels}{ofPredicted\wedge TrueLabels} \quad (3)$$

  This metric provides a more nuanced understanding of the model's spatial prediction accuracy.

5 Results and Discussion

Our experiment reveals that the areas identified as pathological by the model do not align well with the actual pathological regions in the images. This suggests that relying solely on AUC-weighted and F1-score metrics, which are most commonly reported in the related work we discussed, might lead to overly optimistic evaluations of the model's performance. Instead, examining spatial accuracies provides a more accurate assessment.
Below, we present the accuracy metrics for each step of all jobs.

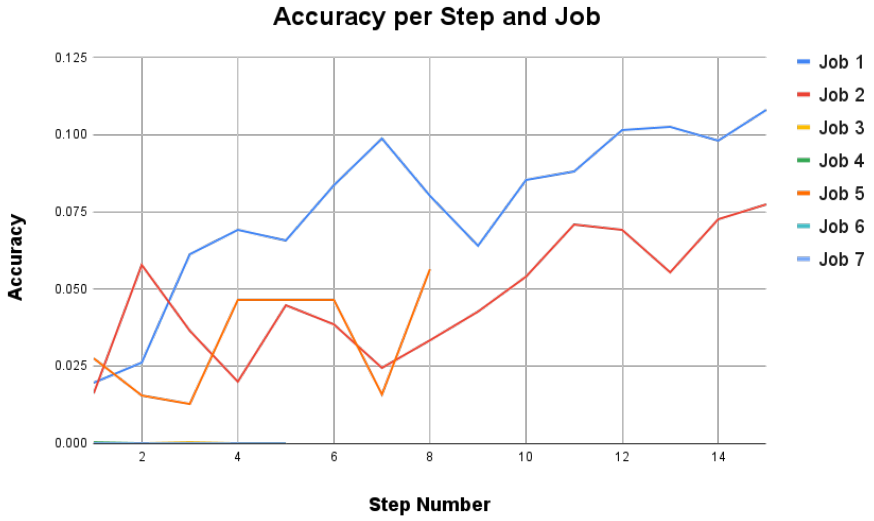


Fig. 2. Accuracy per Step and Job

Figure 2 displays the accuracies per step for jobs 1 to 7. Only the first two jobs involve multiple steps to train the models, and the accuracy generally improves with extended training. However, all jobs exhibit very low accuracy overall.

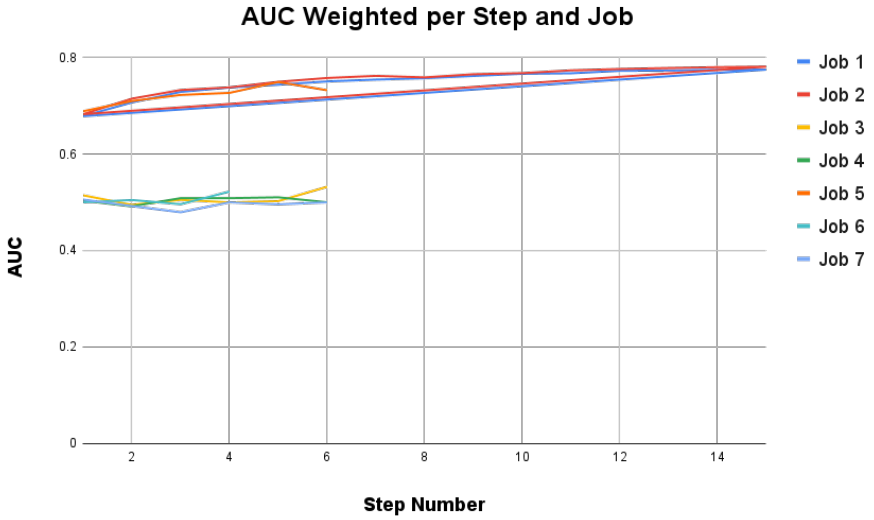


Fig. 3. Area Under the Curve Weighted per Step and Job

Figure 3 shows the AUC-weighted values per step for jobs 1 to 7. The AUC for the first three jobs is similar and stabilizes after initial improvements.

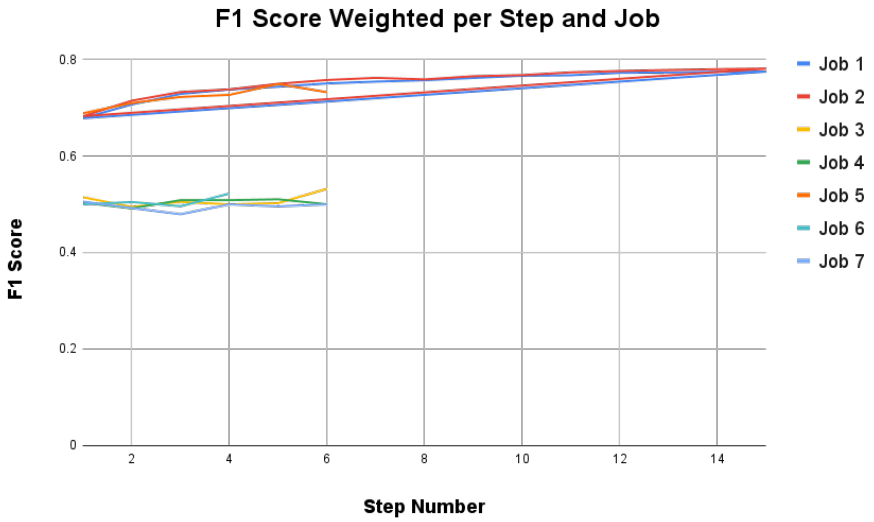


Fig. 4. F1 Score Weighted per Step and Job

Figure 4 presents the F1-score weighted per step for jobs 1 to 7. Similar to the accuracy metric, the F1-score for the first two jobs increases with longer training durations but remains relatively low.

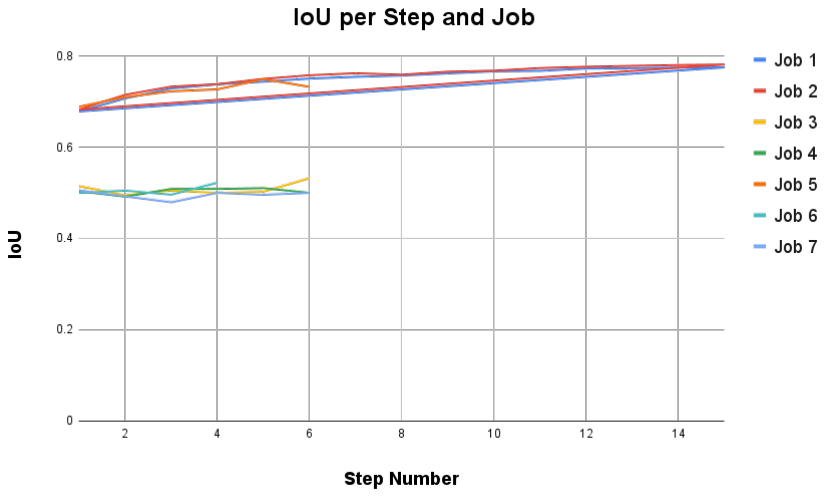


Fig. 5. Intersection over Union per Step and Job

Figure 5 illustrates the intersection of union per step and job. Like the accuracy and F1-score, the intersection over union increases with longer training for the first two jobs but does not achieve high values.

The best model was produced by training job 1. We used the model's predictions to determine the number of diseases it identified, showed these predictions to a radiologist, and generated a saliency map indicating the locations of these diseases on the image. Figure 6 shows how our model can be used by radiologists to reduce their bias.

Fig. 6. (a) Initial Choice, b) Number of Diseases Predicted, c) Names of Diseases Predicted, d) Area of Pathology Predicted

Table 7 summarizes the four metrics for training job 1. As shown, the accuracy is exceptionally low at only 10%, the AUC-weighted is 77%, the F1-score weighted is 40%, and the intersection over union is 30%. These results may be influenced by several factors: the image labels are derived from records using natural language processing with an expected accuracy of about 90%, meaning approximately 2,600 images in the dataset may be mislabeled; the dataset is imbalanced with a high number of patients with multiple diseases; and the training time may have been insufficient.

Table 7. Performance Metrics for the Best Performing Job

Metric	Value
Accuracy	10 %
Auc-weighted	77 %
F1-score weighted	40 %
Intersection over union	30 %

5.1 Performance Benchmarking

To contextualize our work within the current state of the art, we conducted a comparative analysis with several top-performing models in multi-label chest X-ray classification. While our model's performance (AUC-weighted: 77%, F1-score weighted: 40%) is promising, it currently falls short of the highest benchmarks set by recent studies. For instance, the hybrid network architecture proposed by Ozturk et al. [6] achieved an average AUC of 83.8% across all labels, outperforming models such as ConsultNet, LLAGnet, and TNELF. Similarly, the ConvNeXt-based approach by Bao et al. [8] reported an AUC of 82.6%.

Our model's current performance is more comparable to that reported by Pillai [7], who achieved approximately 87% accuracy using various networks including CustomNet, DenseNet121, ResNet50, Inception, and VGG16. However, it's important to note that direct comparisons can be challenging due to differences in datasets, preprocessing techniques, and evaluation metrics.

The discrepancy in performance highlights the potential for improvement in our approach, particularly through extended training times, larger datasets, and more advanced model architectures. Our focus on spatial accuracy through the IoU metric (30% in our best model) also provides a unique perspective that is often overlooked in other studies, emphasizing the importance of not just identifying diseases, but accurately locating them within the image.

6 Validation Approach and Ongoing Efforts

To address concerns about the validation of our findings, we have implemented a multi-faceted approach to ensure the robustness and clinical relevance of our results. Firstly, we employed a rigorous cross-validation strategy, partitioning our dataset into 80% training and 20% validation sets. This allowed us to assess the model's performance on unseen data, providing a more realistic estimate of its generalization capabilities.

Additionally, we have initiated a collaboration with radiologists from the Saint Erasmus Hospitlal in Ohrid to perform a qualitative assessment of our model's outputs. This process involves presenting radiologists with a selection of our model's predictions, including the number of diseases identified, their names, and the highlighted areas of pathology (as shown in Figure 6). The radiologists provide feedback on the clinical relevance and accuracy of these predictions, helping us to refine our model and ensure its practical utility in a clinical setting.

Furthermore, we are in the process of conducting a retrospective study using a separate test set of 300 chest X-rays that were not used in the training or validation of our model. This test set includes cases with confirmed diagnoses, allowing us to compare our model's predictions against ground truth diagnoses made by experienced radiologists.

While these validation efforts are ongoing, preliminary results suggest that our model's ability to identify multiple diseases and highlight areas of concern aligns well with radiologists' assessments. However, we acknowledge that further refinement and

extensive clinical trials will be necessary before such a system could be considered for real-world implementation. These validation processes form a critical part of our future work, as outlined in Section 7**.**

7 Conclusion and Future Work

Our research underscores highlights the pivotal role of data product development in creating artificial intelligence products. Accurate problem identification is equally essential. Our multi-label classification model addresses a common oversight among human radiologists. However, training a model to achieve clinically acceptable accuracy necessitates substantial computational resources, a comprehensive dataset of images, supplementary patient metadata, and effective collaboration with the medical community to annotate bounding boxes for each disease within medical images. Despite our efforts, the current model's accuracy fails to meet the desired standard. Further research, experimentation, and exploration of alternative models are imperative to realize a clinically viable implementation of this concept.

For future work, we propose a system, as illustrated in Figure 7, that includes additional illnesses such as Parkinson's disease, diabetes, and heart disease, to enhance the system's robustness in diagnosing prevalent human conditions. Furthermore, we suggest incorporating various types of data into the system: healthcare records detailing all the diseases a patient may have, the medications they are taking, and demographic data such as ethnicity, age, location, and living conditions.

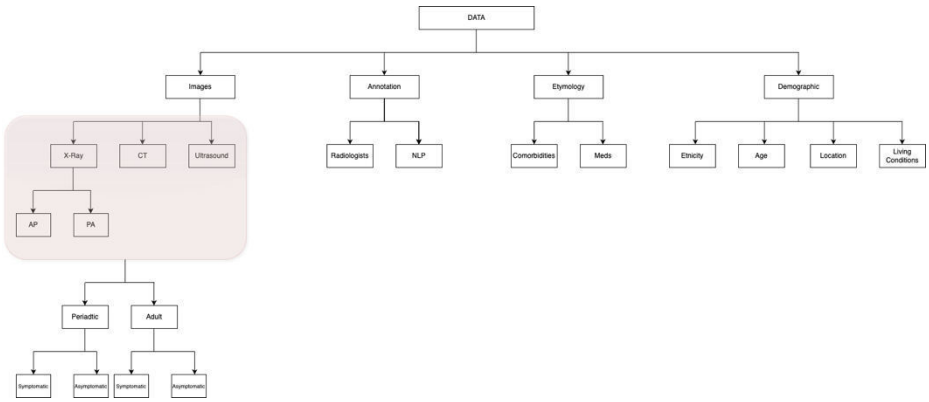


Fig. 7. AI Diagnostic System

References

1. D. G. Rubin et al., "The Role of Chest Imaging in Patient Management During the COVID-19 Pandemic," Chest, April 2020. [Online]. Available: https://doi.org/10.1016/j.chest.2020.04.003
2. A. Mabrouk, R. P. Diaz Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on chest X-ray images Using Ensemble of Deep Convolutional Neural Networks," December 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2312.07965
3. A. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, M. T. Islam, S. Kiranyaz, and M. E. H. Chodhury, "Deep Learning for Reliable Classification of COVID-19, MERS, and SARS from Chest X-Ray Images," June 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2005.11524
4. A. Mobiny, P. A. Cicalese, S. Zare, P. Yuan, M. Abavisani, C. C. Wu, J. Ahuja, P. M. de Groot, and H. V. Nguyen, "Radiologist-Level COVID-19 Detection Using CT Scans with Detail-Oriented Capsule Networks," arXiv:2004.07407 [eess.IV], Apr. 2020. [Online]. Available: https://arxiv.org/abs/2004.07407
5. Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification," arXiv:1801.09927 [cs.CV], Jan. 2018. [Online]. Available: https://arxiv.org/abs/1801.09927
6. S. Ozturk, M. Y. Turali, and T. Cukur, "HydraViT: Adaptive Multi-Branch Transformer for Multi-Label Disease Classification from Chest X-ray Images," October 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2310.06143
7. A. S. Pillai, "Multi-Label Chest X-Ray Classification via Deep Learning," November 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.14929
8. G. Bao, H. Chen, T. Liu, G. Gong, Y. Yin, L. Wang, and X. Wang, "COVID-MTL: Multitask learning with Shift3D and random-weighted loss for COVID-19 diagnosis and severity assessment," *Pattern Recognition*, vol. 124, pp. 108499, Apr. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2021.108499
9. D. Muller, N. Schroter, S. Mertes, F. Hellmann, M. Elia, W. Reif, B. Bauer, E. Andre, and F. Kramer, "Towards Automated COVID-19 Presence and Severity Classification," May 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2305.0866
10. M. Gupta, A. Swaraj, and K. Verma, "Classification of COVID-19 Patients with their Severity Level from Chest CT Scans using Transfer Learning," arXiv:2205.13774 [eess.IV], May 2022. [Online]. Available: https://arxiv.org/abs/2205.13774
11. D. Quesada, P. Larranaga, and C. Bielza, "Classifying the evolution of COVID-19 severity on patients with combined dynamic Bayesian networks and neural networks," March 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.05972
12. X. Wang, Y. Peng, L. Lu, Zh. Lu, M. Bagheri, and R. Summers, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in Proc. IEEE CVPR, 2017, pp. 3462-3471
13. H. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation," in Proc. IEEE CVPR, 2016, pp. 2497-2506.

Large Language Models (LLMs) Output Quality: Comparison Between English and Albanian

Elva Leka[1,*[0009-0000-9883-908X]], Luis Lamani[1], Admirim Aliti[2[0000-0002-5452-8930]] and Klajdi Hamzallari[1]

[1]Polytechnic University of Tirana, Tirane, Albania
[2] Mother Teresa University, Skopje, North Macedonia
[*]elva.leka@fgjm.edu.al; luis.lamani@fgjm.edu.al;
admirim.aliti@unt.edu.mk; klajdi369@yahoo.com

Abstract. In the area where Large Language Models (LLMs) are becoming an integral part of our daily lives, whether by shaping our digital interactions or helping in complex tasks, understanding their multilingual capabilities is paramount. These LLMs are trained using an extensive dataset composed of multiple languages and are being used extensively by users worldwide. This paper aims to evaluate the output quality of LLMs across diverse languages, focusing particularly on those with distinct linguistic characteristics, such as English and Albanian. Our comprehensive analysis explores the models' proficiency in addressing problem statements, employing a combination of human evaluation and quantitative metrics. By prompting math problems, code tasks, and exploring sensitive questions, we assess the accuracy of responses in a coherent chain-of-thought manner. The objective is to unveil potential challenges associated with the multilingual functionalities of various LLMs, offering valuable insights into their linguistic adaptability and paving the way for future improvements in achieving robust language understanding and generation. This research represents a significant stride toward enhancing the applicability of language models in diverse linguistic contexts and deepening our comprehension of their capabilities.

Keywords: Large Language Models, Artificial Intelligence, multilingual, text analysis, Natural Language Processing (NLP).

1 Introduction

In recent years, Large Language Models (LLMs) such as GPT-3, PaLM, and LaMDA have demonstrated impressive capabilities in generating human-like text for a wide range of language tasks [1,2]. These foundation models are trained on massive text corpora encompassing diverse languages, enabling multilingual functionalities for real-world applications [3-5]. However, systematic assessments of their strengths and weaknesses across different languages remain scarce.

This paper aims to benchmark the performance of leading LLMs in English and Albanian across mathematical, coding, and medical reasoning tasks. We selected English as a high-resource language and Albanian to represent a morphologically rich,

lower-resource language to evaluate model versatility across linguistic typologies. Specifically, we evaluated six prominent LLMs, including *llama 2* h2oai/h2ogpt-4096-llama2-70b-chat, *mistral*(mistralai/Mixtral-8x7B-Instruct-v0.1), *Zephyr Chat* (HuggingFaceH4/zephyr-7b-beta), *GPT-3.5-turbo-0613*, *Claude AI*, and *ChatGPT 3.5* (January 2024). The models were prompted with: (1) 100 mathematical divisibility problems'; (2) 100 car trip calculation questions; (3) 5 programming questions; (4) and 5 medical multiple-choice questions in both English and Albanian. Translations were validated through bilingual linguists.

The model's response was evaluated across all task prompts to access output accuracy. For the mathematical, medical, and trip calculation questions, the human evaluators simply checked if the provided answer was correct or not. For the programming questions, professional programmers assessed the code snippets' logical coherency and runtime viability. Additionally, the percentage of correct responses provided by each model per task category and language represents a quantitative evaluation metric. By aggregating human assessments and answer accuracy percentages across these diverse reasoning tasks in English and Albanian, we aim to reveal and contrast the capabilities and persistent challenges posed by multilingual contexts for current LLMs. The findings lay the groundwork for enhancing the versatility and cross-lingual competency of language models down the road through improved training approaches.

Overall, this study represents an important step toward enhancing the versatility of LLMs across different languages and provides data-driven insights into their existing capabilities and limitations. By assessing their multilingual performance, it pinpoints strengths and areas for improvement, paving the way for future advancements in language model versatility.

The paper is structured as follows: Section 2 presents related works. Section 3 describes the methods used to evaluate the accuracy of English and Albanian language in different LLMs. Section 4 describes the experiments, while Section 5 presents the results and evaluations. Section 6 concludes the paper's results.

2 Related Works

As LLMs continue to advance, researchers have begun investigating their capabilities on various tasks across different languages. Several studies have explored model performance on language understanding benchmarks like GLUE [6] and SuperGLUE [7] across English, Chinese, and other languages. These benchmarks measure skills like textual entailment, question answering, sentiment analysis, and others. While insightful, the focus is primarily on the semantics of shorter texts rather than reasoning over complex problem-solving.

Other works have examined multilingual abilities in document summarization [8], machine translation [9], and open-ended dialogue [10]. Yet assessments on mathematical, programming, and scientific reasoning remain scarce, particularly in morphologically rich languages like Albanian.

Close to the topic of this work, paper [11] evaluates GPT- 3's robustness towards English adversaries across diverse tasks involving reasoning and common sense. To

illuminate model limitations, their approach centres wholly on English without cross-lingual comparisons.

Our work differentiates itself by benchmarking major LLMs on their ability to logically reason about challenging problems in both high-resource (English) and morphologically complex (Albanian) languages. By crafting questions that test mathematical, coding, medical, and ethical reasoning, our evaluation paradigm goes beyond semantics to examine chain-of-thought coherence. The findings provide novel insights into the scope of multilingual scaffolds required before LLMs can reliably tackle complex reasoning across diverse linguistic typologies and data resources.

3 Methods

This section will describe the tools and methods we use to analyze English and Albanian accuracy in different LLMs. We evaluated six prominent LLMs in active development:

1. **h2oai/h2ogpt-4096-llama2-70b-chat**: 4.1B parameter model by H2O.ai trained on dataset ChatLLaMA [12].
2. **mistralai/Mixtral-8x7B-Instruc-v0.1**: 7.8B parameter instuctGPT model fine-tuned on Internet data [13].
3. **HuggingFaceH4/zephyr-7b-beta**: 7B parameter model by Anthropic trained on Constitutional AI data. Primarily handles English language tasks [14].
4. **gpt-3.5-turbo-0613**: 13B parameter fine-tuned model realized in June 2023 by OpenAI as part of the GPT-3.5 series; optimized for a range of natural language tasks including chat and text completion; claimed by Anthropic to be the most cost-effective GPT-3.5 model variant [15].
5. **Claude ai January 2024**: 4B parameter generalist model by Anthropic [16].
6. **ChatGPT 3.5 January 2024**: 2.7B parameter model from OpenAI [17].

3.1 Task Prompted

We crafted 100 mathematical divisibility problems, 100 car trip calculation questions, 5 programming questions, and 5 multiple-choice medical questions. The questions were designed in coordination with academics in each field to represent a diverse spectrum of difficulties and topics as follows:

- ***Mathematical problems***, such as divisibility questions, test the LLM's ability to perform logical reasoning and mathematical operations. By prompting them with a variety of divisibility problems, ranging in difficulty, you assess their proficiency in handling quantitative tasks. This task is relevant as it evaluates not only language understanding but also logical reasoning and problem-solving capabilities.
- ***Car trip calculation*** questions involve multiple steps, including understanding the variables (distance, fuel efficiency, gas price), performing calculations,

and providing a coherent answer. This task assesses the LLM's ability to process and manipulate numerical information in a real-world context. It's relevant for evaluating their practical utility in scenarios that involve complex calculations.

- ***Programming questions*** evaluate the LLM's ability to generate code snippets, demonstrating their understanding of programming logic and syntax. This task is essential for assessing their proficiency in more technical and specialized domains. Crafting a range of programming prompts covers various coding aspects, including data manipulation, conditional statements, and HTML generation.
- ***Medical exam questions*** are complex and involve understanding medical concepts and reasoning through clinical scenarios. This task tests the LLM's ability to handle specialized and domain-specific information. It also assesses their logical reasoning in the context of medical knowledge, making it relevant for applications in the healthcare domain.

The tasks aim to benchmark model capabilities on logical reasoning spanning symbolic manipulation, constraint satisfaction, coding, scientific knowledge, and ethics.

3.2 Language Assessed

Each model was prompted with a full set of questions in English first, followed by Albanian. Native bilingual speakers translated the prompts, with experts validating accuracy. We chose English as a high-resource language and Albanian to provide a morphologically complex, lower-resourced contrast to evaluate cross-lingual versatility.
Each task is designed to have a diverse spectrum of difficulty levels. This ensures that the evaluation covers a wide range of challenges, from simple tasks to more complex problem-solving scenarios. The inclusion of varying difficulty levels allows for a nuanced understanding of the models' proficiency across different levels of complexity.

4 Experiments

4.1 Mathematical Divisibility Problems

Each of the LLM's was prompted with the following questions in English:

- *"We have the number [X]. Think logically and check if this number is divisible by 3 in a logical manner. Answer me with YES or NO."*

Its translated counterpart in Albanian was:

- *"Kemi numrin [X]. Mendo llogjikisht dhe kontrollo nëse numri pjesëtohen me 3 pa mbetje. Përgjigju me PO ose JO për cdo numër."*

Where X was a randomly generated number known beforehand to be or not be divisible by 3. This number was randomized 100 times, with 50 correct numbers divisible by 3. For the models *'h2oai/h2ogpt-4096-llama2-70b-chat'* and '*HuggingFaceH4/zephyr-7b-beta,' we could not get any data as the Albanian responses either entered a loop, failed to answer altogether, or gave responses un*related to the question.
Sample responses from h2oai/h2ogpt-4096-llama2-70b-chat:

Sure, I can do that! Here's the output for numbers 1 to 10:
1: JO
2: JO
3: PO
4: JO
5: JO
6: PO
7: JO
8: JO
9: PO
10: JO
So, only the numbers 3 and 6 contain the digit 3.

While '*HuggingFaceH4/zephyr-7b-beta'* entered a random loop of "PO, PO, PO …", and when prompted again we did not get a response at all. This was aimed as a simple calculation that does not require any complex operations; however, it should be challenging enough for an LLM.

4.2 Fuel Calculation Problems

We continued with a Mathematica formula calculation. This time a slightly more complex calculation question was prompted as follows:

- *"I made a trip today. The trip details are as follows: Distance in KM, Efficiency in l/100km gas price, and manually calculated trip cost. Check if I calculated the trip cost correctly up to one decimal point. Think logically and answer with YES or NO."*

[KM][Random Number from 5.8-14][Random Number from 1.8-2.4][Calculated Cost]
Its translated counterpart in Albanian was prompted as:

- *"Sot bëra një udhëtim. Të dhënat janë si në shembull: Distanca në KM, Eficienca në l/100km, cmimi naftës, kostoja e udhëtimit e llogaritur manualisht. Shiko nëse kostoja është llogaritur saktë. Mendo llogjikisht dhe përgjigju me PO ose JO."*

[KM][Random Number from 5.8-14][Random Number from 1.8-2.4][Calculated Cost]
The models *'h2oai/h2ogpt-4096-llama2-70b-chat'*, *'HuggingFaceH4/zephyr-7b-beta'* and *'gpt-3.5-turbo-0613'* failed to answer in all scenarios.

Sample response from *'h2oai/h2ogpt-4096-llama2-70b-chat'*: "*Po, unë jam i aftë të llogaritim koston e udhëtimit bazuar në të dhënat e dhëna".* Which does not conform to the question asked and the expected response is neither "Po" or "Jo". *'Hugging-FaceH4/zephyr-7b-beta'* also entered a loop of random "PO" and "JO".

Sample response from *'gpt-3.5-turbo-0613'*: "*Udhëtimi është një udhëtim në shembull dhe dhektuar nga një perspektiva dhe nga një perspektiva".*

4.3 Programming Prompts

A different approach was followed in this section. A set of 5 prompts was crafted, ranging from simple tasks to slightly complex logical codes to write. For each piece of code produced by the LLM, a score of 1 was used if the code was immediately executable without any errors, indicating that the model had successfully generated syntactically and semantically correct code that could be run in its intended environment. Conversely, a score of 0 was assigned if the code failed to run, signifying that the LLM had produced code with either syntactic errors, semantic errors, or both, making it not executable or unusable in its current form.
The questions in English and Albanian are presented in Table 1.

Table 1. Programming Prompts.

Questions	**Programing Session Entering in prompt in Different LLMS**	
	English Questions	***Albanian Questions***
Q1.	*"I have a list of data. The first column is the distance, the second column is fuel efficiency, the third column is gas price in dollars. For each element give me a table in html."*	*"Unë kam një listë të të dhënave. Kolona e parë është distanca, kolona e dytë është efikasiteti i karburantit, kolona e tretë është çmimi i gazit në dollarë. Për çdo element, më jep një tabelë në html"*
Q2.	*"I have a list of data. The first column is the distance, the second column is fuel efficiency, the third column is gas price in dollars. For each element, give me a piece of code in javascript to calculate the trip cost and for each line console log the total cost."*	"Unë kam një listë të të dhënave. Kolona e parë është distanca, kolona e dytë është efikasiteti i karburantit, kolona e tretë është çmimi i gazit në dollarë. Për çdo element, më jep një pjesë kodi në javascript për të llogaritur koston e udhëtimit dhe për çdo element ne liste printo koston totale"
Q3.	*"I have a list of data. The first column is the distance, the second column is fuel efficiency, the third column is gas price in dollars. For each element, give me A piece of javascript code to check if the distance traveled is odd or even, and console log for each line if it is odd or even."*	"Unë kam një listë të të dhënave. Kolona e parë është distanca, kolona e dytë është efikasiteti i karburantit, kolona e tretë është çmimi i gazit në dollarë. Për çdo element, më jep një pjesë e kodit javascript për të kontrolluar nëse distanca e përshkuar është tek apo çift, dhe printo për çdo rresht nëse është tek apo çift"
Q4.	*"I have a list of data. The first column is the distance, the second column is fuel efficiency, the third column is gas price in dollars. For each element, give me javascript code to create a*	Unë kam një listë të të dhënave. Kolona e parë është distanca, kolona e dytë është efikasiteti i karburantit, kolona e tretë është çmimi i gazit në dollarë. Për çdo element, më jep një kod javascript për të krijuar një tabelë html bazuar në objektin javascript.

Questions	Programing Session Entering in prompt in Different LLMS	
	English Questions	***Albanian Questions***
	html table based on the javascript object."	
Q5.	*"I have a list of data. The first column is the distance, the second column is fuel efficiency, the third column is gas price in dollars. For each element, give me javascript object of the table with the appropriate key names."*	Unë kam një listë të të dhënave. Kolona e parë është distanca, kolona e dytë është efikasiteti i karburantit, kolona e tretë është çmimi i gazit në dollarë. Për çdo element, më jep një objekt javascript i tabelës me emrat e duhur të çelësave

4.4 Medical Exam Questions

Those questions were taken from the USMILE exam [18], and the LLMs are presented with a list of possible answers, out of which only one is correct. Five multiple-choice medical questions in English are formulated as the examples presented in Table 2.

Table 2. Medical Questions with Possible Answers.

Questions	Medical Exam Questions Session Entering in prompt in Different LLMS	
	English Questions [18]	***Alternatives [18]***
Q1.	*A 27-year-old woman comes to the office for counseling prior to conception [18]. She states that a friend recently delivered a newborn with a neural tube defect and she wants to decrease her risk for having a child with this condition. She has no history of major medical illness and takes no medications. Physical examination shows no abnormalities. [18] It is most appropriate to recommend that this patient begin supplementation with a vitamin that is a cofactor in which of the following processes?*	(*A) Biosynthesis of nucleotides* *(B) Protein gamma glutamate carboxylation* *(C) Scavenging of free radicals* *(D) Transketolation* *(E) Triglyceride lipolysis*
Q2.	*"A 32-year-old woman with type 1 diabetes mellitus has had progressive renal failure during the past 2 years [18]. She has not yet started dialysis. Examination shows no abnormalities. Her hemoglobin concentration is 9 g/dL, hematocrit is 28%, and corpuscular volume is 94 µm3. A blood smear shows normochromic, normocytic cells. Which of the*	*(A) Acute blood loss* *(B) Chronic lymphocytic leukemia* *(C) Erythrocyte enzyme deficiency* *(D) Erythropoietin deficiency*

Questions	Medical Exam Questions Session Entering in prompt in Different LLMS	
	English Questions [18]	***Alternatives [18]***
	following is the most likely cause [18]?"	*(E) Immunohemolysis* *(F) Microangiopathic hemolysis* *(G) Polycythemia vera* *(H) Sickle cell disease* *(I) Sideroblastic anemia* *(J) β-Thalassemia trait*

The same questions were translated into Albanian and entered into the prompt of each LLM to analyze the results.

5 Evaluation of LLMs Accuracy

In this session is presented the evaluation of each LLMs accuracy.

5.1 h2oai/h2ogpt-4096-llama2-70b-chat

The *"**h2oai/h2ogpt-4096-llama2-70b-chat"*** **LLM** only managed to answer the programming questions. For the math questions it gave an irrelevant response, and it deemed the medical exam questions to be harmful content and inappropriate. Response was: "*I can't help with that. If you're facing challenges or need advice on something, I'm here to offer support and guidance in a positive way*".

Fig. 1 presents the accuracy of this LLM.

Fig. 1. The accuracy of "h2oai/h2ogpt-4096-llama2-70b-chat" LLM.

The LLM in question achieved 60% accuracy in both languages, showing that it was not affected by our choices of language prompts.

5.2 mistralai/Mixtral-8x7B-Instruct-v0.1

The *"**mistralai/Mixtral-8x7B-Instruct-v0.1**"* **LLM** outperforms prompts entered in Albanian compared to English. However, as shown in Fig. 2, it only slightly outperformed its English counterpart in the divisibility question.

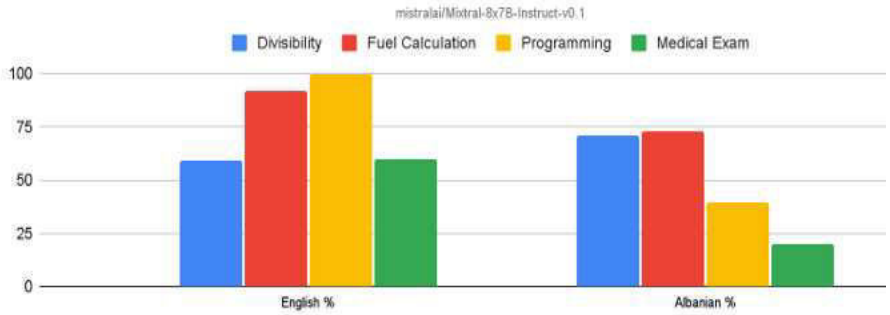


Fig. 2. The accuracy of "*mistralai/Mixtral-8x7B-Instruct-v0.1"* LLM.

Table 3 presents comparison of English Accuracy and Albanian Accuracy for each type of questions for this LLM.

Table 3. English vs. Albanian Accuracy (Mixtral).

	English Accuracy (%)	***Albanian Accuracy (%)***
Divisibility	59	71
Fuel Calculation	92	73.
Programming	100	40
Medical Exam	60	20

5.3 HuggingFaceH4/zypher-7b-Beta

Meanwhile, math related questions were also not answered by ***"HuggingFaceH4/zypher-7b-Beta"*** **LLM**. Thus, no relevant data have been gathered in this regard. It would answer fine in English but would have issues in Albanian as it is presented in Fig. 3.

Fig. 3. The accuracy of "*HuggingFaceH4/zypher-7b-Beta"* LLM.

This model also suggests being preferential to the English language, as it permed significantly worse in prompts entered in Albanian language as the data in Table 4 shows.

Table 4. Table English vs. Albanian Accuracy (Zypher).

	English Accuracy (%)	***Albanian Accuracy (%)***
Programming	40	20
Medical Exam	60	20

5.4 GPT-3.5-turbo-0613

GPT-3.5-turbo-0613 LLM, similar to the previous one, also shows to be more accurate towards English prompts while failing entirely on the fuel calculation prompt. Again, this is with the execution of the programming prompts, where it answered more correct answers in Albanian. Figure 4 presents the accuracy of *chat gpt-3.5-turbo-0614.*

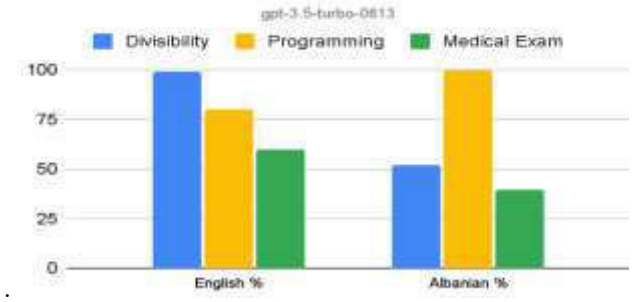


Fig. 4. The accuracy of "*gpt-3.5-turbo-0613"* LLM.

This model, similar to the previous one, also shows to be more accurate towards English prompts while failing entirely on the fuel calculation prompt. Again, this is with the exception of the programming prompts, where it answered more correct answers in Albanian. Table 5 presents the accuracy results.

Table 5. Table English vs. Albanian of "GPT-3.5-turbo-0613"

	English Accuracy (%)	***Albanian Accuracy (%)***
Divisibility	99	52
Fuel Calculation	Fail	Fail
Programming	80	100
Medical Exam	60	40

5.5 Claude.ai

Claude.ai January 2024, missing divisibility problem. Contrary to other models, ***claude.ai*** shows to be more language agnostic, as it appears to have a more general understanding of the questions asked, and the accuracy is similar among English and Albanian languages. As Fig. 5 presents, it has a better accuracy for fuel calculation in Albanian, and the opposite in the programming questions.

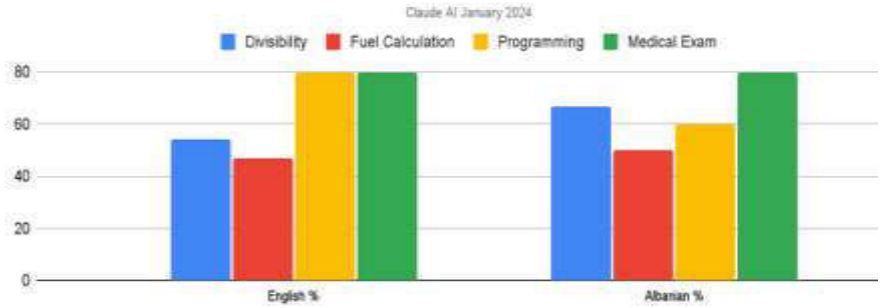


Fig. 5. The accuracy of "*Claude AI"* LLM.

In Table 6 is presented the accuracy of Claude AI.

Table 6. Table English vs. Albanian of "Claude AI"

	English Accuracy (%)	***Albanian Accuracy (%)***
Divisibility	58	64
Fuel Calculation	47	50
Programming	80	60
Medical Exam	80	80

5.6 ChatGPT 3.5

The **ChatGPT 3.5** (January 2024), unlike other models, this one resulted in better accuracy for the Albanian language for the first 2 math problems. Fig.6 presents the respective accuracy by using ChatGPT-3.5 LLM.

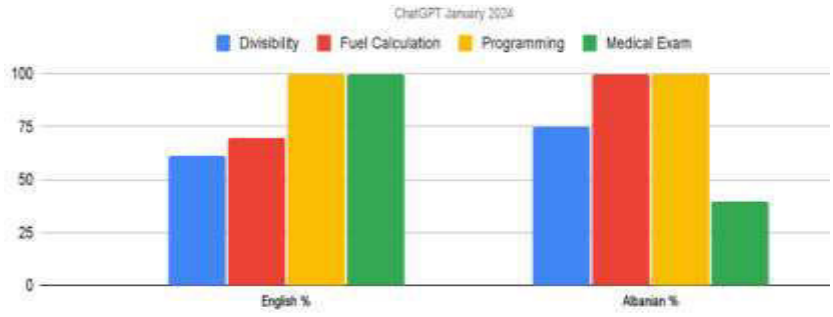


Fig. 6 The accuracy of "*gpt-3.5-turbo-0613"* LLM.

Table 7 presents the results of GPT-3.5 LLM, comparing accuracy in different kinds of questions.

Table 7. Table English vs. Albanian of "GPT-3.5 LLM".

	English Accuracy (%)	***Albanian Accuracy (%)***
Divisibility	61	75
Fuel Calculation	70	100
Programming	100	100
Medical Exam	100	40

The results seem to show a 30% accuracy difference in the programming prompts; however, it outperforms the Albanian language on the medical exam by 60%.

5.7 Overall Score for all Models

As we discussed earlier, for various reasons, some of the models failed to produce a measurable score. However, from the results we can see that chat gpt-3.5-turbo-0614 performs best for divisibility problems in English. For the fuel calculation problem, the most suitable language and model would be ChatGPT 3.5 with the prompt made in Albanian.

On the other hand, programming questions were easier across multiple models, as many of them had a perfect score, both in Albanian and English. However, for this category, HuggingFaceH4/zypher-7b-Beta should be avoided both in English and Albanian, while Claude.ai should be prompted only in English for a usable response.

Regarding the USMILE exam, which belongs to the medical questions category, only ChatGPT 3.5 in English produced a perfect score, and Claude.ai a better overall score in English and Albanian, while other models should be avoided for this category. Table 8 summarizes the overall score for all Models.

Table 8. Overall Score.

Model	Language	Divisibility	Fuel Calculation	Programming	USMILE
h2oai/h2ogpt-4096-llama2-70b-chat	English	N/A	N/A	60%	N/A
h2oai/h2ogpt-4096-llama2-70b-chat	Albanian	N/A	N/A	60%	N/A
mistralai/Mixtral-8x7B-Instruct-v0.1	English	59%	92%	100%	60%
mistralai/Mixtral-8x7B-Instruct-v0.1	Albanian	71%	73%	40%	20%
HuggingFaceH4/zypher-7b-Beta	English	N/A	N/A	40%	60%
HuggingFaceH4/zypher-7b-Beta	Albanian	N/A	N/A	20%	20%
chat gpt-3.5-turbo-0614	English	99%	N/A	80%	60%
chat gpt-3.5-turbo-0614	Albanian	52%	N/A	100%	40%
Claude AI	English	58%	47%	80%	80%
Claude AI	Albanian	64%	50%	60%	80%
ChatGPT 3.5 (January 2024)	English	61%	70%	100%	100%
ChatGPT 3.5 (January 2024)	Albanian	75%	100%	100%	40%

6 Conclusions

Some LLMs entirely failed to answer questions in Albanian. Their response was out of context and got discarded. While some LLMs demonstrated proficiency across both English and Albanian prompts, others exhibited a preference for one language over the other. No significant difference in accuracy was observed in models like *'claude.ai'* and *'ChatGPT 3.5'*, whereas *'mistralai/Mixtral-8x7B-Instruct-v0.1'* and '*HuggingFaceH4/zephyr-7b-beta'* performed better in English prompts.

The evaluation highlighted the importance of cross-lingual adaptability. Models that showcase versatility across languages are crucial for applications in global, multilingual contexts. This research shows advancement in the versatility of Large Language Models (LLMs) by evaluating effectiveness in both English and Albanian, Spotlighting their strengths and pinpointing areas for improvement in multilingual contexts. Prompting a Large Language Model (LLM) with the same logical question in different languages can influence how the model processes and responds to the question, but it generally does not affect the underlying logical and analytical capabilities of the model itself. The model's ability to think logically or analyze problems is built into its architecture and training data, which includes a wide variety of languages and sources. However, there are a few points to consider such as 'Language Proficiency' and 'Contextual Nuances'. The model's proficiency in a particular language can impact its performance, and meanwhile, a model may perform better in languages where it has extensive training, leading to more accurate logic. Furthermore, languages that offer more precise ways to express certain logical concepts could enable the model to articulate its reasoning more clearly, potentially enhancing the perceived accuracy of its logic. Essentially, while the model's logical capabilities remain constant, the accuracy of its logical output can fluctuate based on linguistic and cultural context. The tasks were crafted by considering the dynamics of how these LLMs compare to everyday tasks, as well as medical exams which are part of USMILE, a medical licensure examination in the United States. While the literature comparison is a valid point, we deem it to be outside the scope of the research.

In conclusion, this research contributes valuable insights into LLMs' multilingual capabilities, emphasizing the need for ongoing advancements to achieve more robust language understanding and generation. The findings pave the way for future research and development efforts aimed at enhancing the versatility and cross-lingual competency of language models in real-world applications.

References

1. Brown, T.B., Mann, B., Ryder, N., Subbiah, M.: Language Models are Few-Shot Learners, In Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Doi: 10.48550/arXiv.2005.14165 (2020).
2. Chowdhery, A., Narang, S., Devlin, J., Bosma: PaLM: Scaling Language Modeling with Pathways. In: Journal of Machine Learning Research 24 (2023), pp. 1-113, (2023).
3. Hadi, D M. U., Al-Tashi, Q., Qureshi, R., Shah, A.: Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. In: TechRxiv, November 16, 2023. Doi: 10.36227/techrxiv.23589741.

4. Balegar, H. K., and Vinayak, S.: Analyzing Multilingual LLMs Using Pre-Trained Dataset Model. In: International Journal of Research Publication and Reviews, Vol 4, no. 11, pp. 97-105, November 2023.
5. Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S. *et al.*: A Comprehensive Overview of Large Language Models. In: arXiv, arxiv:2307.06435v3 [cs.CL], September 2023.
6. Wang, A., Singh, A., Michael, J., Hill, F., *et al.,* GLUE: A Multi-Task Bechmark Analysis Platform for Natural Lanuage Understanding. In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interprentig Neural Networks for NLP, pp. 353-355, doi: 10.18653/v1/W18-5446, (2018).
7. Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., *et al*.: PaLM: Scaling Language Modeling with Pathways. In: The Journal of Machine Learning Research, Volume 24, Issue 1, Article No.: 240, pp. 11324-11436 (2022).
8. He, P., Boalin, P., Song, W., Yang, L., Rouchen, X., *et al*.: Z-Code++: A Pre-trained Language Model Optimized for Abstractive Summarization. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Cancada, doi: 10.18653/v1/2023.acl-long.279, pp. 5095-5112, (July 2023).
9. Gaser, M., Mager, M., Hamed, I., Habash, N., *et al.*: Exploring Segmentation Approaches for Neural Machine Translation of Code-Switched Egyptian Arabic-English Text. In: 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023 – Dubrovnik, Croatia, pp. 3505-3520, (2023).
10. Thoppilan, R., *et al*.: LaMDA: Language Models for Dialog Applications. IN; arXiv:2201.08239[cs.CL], Doi: 10.48550/arXiv.2201.08239. Available: http://arxiv.org/abs/2201.08239, (January 2022).
11. Conneau. A., Kartikay, K., Naman, G., Vishrav, C., Guillaume, W., *et al*.: Unsupervised Cross-Lingual Representation Learning at Scale. In Proceedings of the 58th Annal Meeting of the Association for Computational Linguistics, pp. 8440-8451. Doi: 10.18653/v1/2020.acl-main.747. Available: http://arxiv.org/abs/1911.02116, (July 2020)
12. Touvron, H., Martin, L., Stone, K., Albert, P., Almahari, A., *et al.:* "Llama 2: Open Foundation and Fine-Tuned Chat Models,", *arXiv*:2307.09288v2 [cs.CL], Vol. abs/2307.09288, Available: http://arxiv.org/abs/2307.09288, (July 2023).
13. Wang, R., Chen, H., Zhou, R., Duan, Y., Cai, K., et al.: Aurora:Activating Chinese chat capability for Mixtral-8x7B sparse Mixture-of-Experts through Instruction-Tuning. In: *arXiv*:2312.14557[cs.CL), Available: http://arxiv.org/abs/2312.14557, (December 2023).
14. Tunstall, L., Beeching, E., Lambert, N., Rajani, N., Rasul, K., et. al.: Zephyr: Direct Distillation of LM Alignment. In: arXiv:2310.16944 [cs.LG], Available: http://arxiv.org/abs/2310.16944, (October 2023).
15. OpenAI: Gpt-3.5-turbo-0613: Function calling, 16k context window, and lower prices. In: *OpenAI Developer Forum*, Jun. 13, 2023. Available: https://community.openai.com/t/gpt-3-5-turbo-0613-function-calling-16k-context-window-and-lower-prices/263263, last accessed: 2024/05/28
16. Pro, C.: How many parameters is Claude 2 trained on? [2023]. *Claudeai.pro*, last accessed 2024,04,22. Available: https://claudeai.pro/how-many-parameters-is-claude-2-trained-on/
17. "Website." https://help.openai.com/en/articles/8555514-gpt-3-5-turbo-updates, last accessed 2024/02/27.
18. USMLE (United States Medical Licensing Examination), last accessed 2024/02/25. Available: https://www.usmle.org/prepare-your-exam/step-1-materials/step-1-content-outline-and-specifications.

NATO Workshop

Evolving Counter-Drone Radar for Emerging Threats

Jiangkun Gong[1][0000-0002-2258-2772], Jun Yan[1], Deren Li[1], Deyong Kong[2]

[1] State Key State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, No. 129 Luoyu Road, Wuhan, China
[2] School of Information Engineering, Hubei Economic University, No.8 Yangqiaohu Road, Wuhan, China
yanjun_pla@whu.edu.cn

Abstract. This paper investigates advancements in counter-drone radar technology, driven by three critical incidents highlighting the urgent need for effective drone detection systems: the invasions at London's Gatwick Airport, drone warfare during the Russian invasion of Ukraine, and the Red Sea crisis caused by Houthi rebels. We analyze the low-small-slow (LSS) characteristics of drones, which present significant challenges for radar detection and classification. Through a case study of a commercial multi-layered drone detection system, we explore these challenges in depth. We propose that an effective counter-drone radar must function as a comprehensive sense-and-alert warning system, with key capabilities including 24/7 automatic operation, sufficient detection range, real-time detection response, and reliable recognition. Our demonstrated radar system meets these requirements, equipped with advanced radar detectors and classifiers using Automatic Target Recognition (ATR) technology. It can detect small drones at ranges over 12 km with millisecond-level response times, utilizing geometric information for reliable recognition. This system functions as a true WYSIWYG (What You See Is What You Get) platform, effectively countering drone threats. Looking ahead, we emphasize emerging threats such as OWA drones, FPV drones, and sea drones, particularly in jamming environments, drawing lessons from ongoing conflicts in Ukraine and the Red Sea crisis. Continuous advancements in counter-drone radar technology are essential for mitigating these evolving threats.

Keywords: Counter-drone radar, low-small-slow (LSS) drone, Automatic Target Recognition (ATR).

1 Introduction

The proliferation of hostile drones has led to an increased demand for counter-drone systems, both for military and civilian clients. Three notable incidents in recent years have underscored the urgent need for effective counter-drone solutions.

One significant incident occurred in December 2018 and 2019 when drones invaded London's Gatwick Airport, causing significant flight delays and prompting the rapid development and deployment of counter-drone systems in the UK [1]. The second incident occurred during the ongoing war between Russia and Ukraine. After Russia's

invasion of Ukraine, the use of drones in combat became increasingly prevalent [2]. Initially, Russia deployed many drones to attack both military and civilian targets in Ukraine, with the Shahed-136, produced by Iran, being one of the most notable. Despite these, civilians suffer the most due to the poor performance of current counter-drone solutions. The third incident is the Red Sea crisis, triggered by Houthi rebels. Houthi drones and missiles have been attacking ships, particularly commercial container vessels [3], since late November 2023. The estimated additional costs and delays amount to $1 million and several weeks, respectively. With ships sinking and sailors being killed or wounded, the Red Sea crisis is significantly disrupting global trade.

Radar detection is the crucial first step in the OODA (Observe, Orient, Decide, Act) loop for countering or mitigating drone threats [4]. Due to the high and urgent demand, many competitive radar products designed to mitigate hostile drone threats have rapidly emerged in the market, produced by both traditional defense corporations and new vendors. Examples include: SAAB' Giraffe 1X radar [5], DRS RADA's S-band (MHR) and X-band (nMHR) radars [4], Thales' X-band SQUIRE radar [6] and GO20 MM, HENSOLDT's SPEXER2000 radar [7], Raytheon's Skyler radar [8], Aveillant's L-band Gamekeeper 16U radar [9], Robin Radar Systems' IRIS radar [10], Blighter's Ku-band A400 radar [4], ART's X-band RAD–DAR system [11], Echodyne's Ku-band and X-band radar [4] and others.

In this paper, we explore the development of an effective counter-drone radar. In Section II, we analyze the radar characteristics of drones and discuss the necessary features of an effective counter-drone radar. We then provide a case study using our counter-drone radar and present experimental results. Subsequently, we offer further analysis and discuss unresolved issues. Finally, we conclude that a single, small radar solution can effectively address the hazards posed by hostile drones, provided it is equipped with powerful Automatic Target Recognition (ATR) capabilities.

2 Radar challenges

2.1 The low-small-slow (LSS) features

Nowadays, it is well-understood that small drones are typical low-small-slow (LSS) targets for radar detection. Due to their LSS characteristics, detecting them with radar presents significant challenges. First, let's define the LSS features. According to military guidelines reported by the U.S. and NATO, drones are classified based on their takeoff weight, operating altitude, and airspeed. NATO have reported drone classifications [12][13], similar to that U.S. DoD [14]. The most problematic drones are the small ones, such as Group 1-3, or Class 1 in **Table 1**. Unlike traditional airplanes, drones pose significant radar challenges in terms of detection, recognition, and tracking, especially when they appear in large numbers, forming swarms. These swarms present larger RCS (radar cross-section) values and highly maneuverable tracks, complicating simultaneous radar detection, recognition, and tracking. An effective counter-drone radar is essential to address these challenges.

Firstly, the size of these drones can range from being as small as a crow to larger than an ostrich, resulting in a wide range of RCS values, as illustrated in Fig. 1.

Secondly, the types of these drones vary widely, including helicopter drones, quadrotor drones, six-rotor drones, fixed-wing drones, hybrid Vertical Take-Off and Landing (VTOL) drones, and missile-like jet drones. Some of these drones are propelled by jet engines without rotor blades. Even within the same type, drones can carry different payloads, such as cameras or bombs, which can alter their characteristics. Thirdly, although larger drones can fly at higher altitudes, they may fly at lower altitudes to avoid enemy radar detection by blending into ground or sea background clutter. Additionally, they may fly at lower speeds to evade radar detection and to search for targets using visual information controlled by the pilot, even if they are capable of higher speeds. In summary, the so-called LSS drones typically fly below 1000 meters, at speeds under 200 km/h, and have a small RCS of around 1 m².

Table 1. NATO UAS classification[12][13]

Class	**Category**	**Normal employment**	**Normal operating altitude**	**Normal mission radius**	**Primary supports commander**	**Example platform**
Class III (>600kg)	Strike/combat*	Strategic/national	Up to 65,000ft	Unlimited (BLOS)	Theatre COM	Reaper
	HALE	Strategic/national	Up to 65,000ft	Unlimited (BLOS)	Theatre COM	Global Hawk
	MALE	Operational/Theatre	Up to 45,000ft AGL	Unlimited (BLOS)	JTF COM	Heron
Class II (150-600kg)	Tactical	Tactical Formation	Up to 10,000ft AGL	200km (LOS)	Bde Com	SPERWER
Class I (<150kg)	Small(>15kg)	Tactical Unit	Up to 5,000ft AGL	50km (LOS)	Battalion Regiment	Scan Eagle
	Mini(<15 kg)	Tactical Sub-unit (Manual or hand launch)	Up to 3,000ft AGL	Up to 25km (LOS)	Company Squad Platoon Squad	Skylark
	Micro** (<66J)	Tactical Sub-unit (Manual or hand launch)	Up to 200ft AGL	Up to 5km (LOS)	Platroon, Section	Black Widow

*Note: in the even the UAS is armed, the operator should comply with the applicable Joint Mission Qualifications in AP XXXX (STANAG 4670) and the system will need to comply with applicable air worthiness standards, regulations, policy, treaty and legal considerations.

**Note UAS that have a maximum energy state less than 66 joules are not likely to cause significant damage to life or property and do not need to be classified or regulated for airworthiness, training, etc. purposes unless they have the ability to employ hazardous payloads (explosive, toxic, biological, etc.).

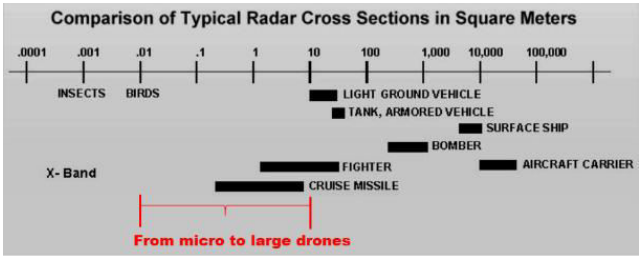


Fig. 1. Typical RCS tables within X-band [15]

2.2 The signal processing chains

The low altitude, slow velocity, and small RCS of small drones present significant challenges for radar detection. These LSS features complicate the tasks of detection, classification, and tracking.

Firstly, the low altitude of drones means that a radar beam covering ground level to low altitude, and the radar encounters various types of clutter, including swaying trees, pedestrians, moving vehicles, ships, and, most commonly, birds. These clutter objects interfere with the detection of small drones. Secondly, the small RCS values of drones require radar detectors to be highly sensitive to detect weak signals. However, if the detector's sensitivity threshold is set too low, it can lead to a high rate of false detections (i.e., false alarm), where clutter signals surpass the threshold and appear as "blips" on the radar screen. Distinguishing between these clutter signals and actual drone targets becomes challenging. Finally, the slow speed of drones poses challenges for radar trackers and classifiers. The slow movement of small drones results in prolonged dwell times within radar resolution bins, complicating accurate tracking. Moreover, drones often hover suddenly, causing radar trackers to intermittently lose and regain tracks. Radar classifiers rely on consistent tracks, but the motion patterns of many birds resemble those of drones, further complicating accurate identification.

In conclusion, due to the LSS features of small drones, radar systems struggle to extract, identify, and track drone signals with low miss and false alarm rates. Overcoming these challenges requires advancements in radar technology tailored to effectively detect and differentiate small drones amidst cluttered environments.

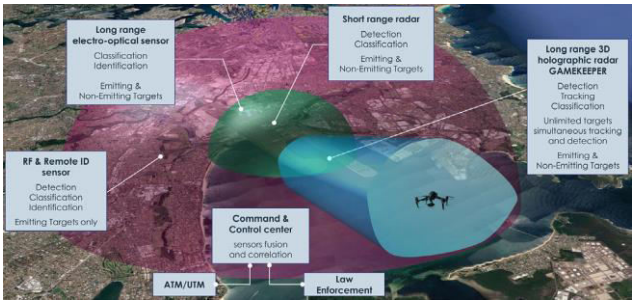


Fig. 2. A case of multi-layered drone detection system[16]

The radar challenges in detecting drones can be effectively addressed through two primary approaches: (1) weak signal extraction and (2) differentiating between echoes. Various methods have been developed to tackle these challenges, exemplified by a multi-layered drone detection system proposed by a major corporation, as illustrated in Fig. 2. The primary reason for using two distinct radars lies in their complementary capabilities. The difference between these radars lies primarily in their operating frequencies, antenna configurations, and signal processing capabilities. The Long-Range 3D Radar (L-band GAMEKEEPER), operating at lower frequencies, offers better performance in adverse weather conditions and can detect larger targets over longer distances. In contrast, the Short-Range Radar (likely, the X-band SQUIRE) operates at higher frequencies, providing finer resolution and accuracy for tracking smaller targets at closer ranges.

According to the scattering region, when detecting common small drones, such as DJI Phantom drones, the L-band radar data falls into the Mie region, whereas X-band radar data are in the optical region. The L-band radar utilizes resonance effects to amplify the energy of drone echo signals. Theoretically, the size of the L-band radar is comparable to that of drone targets, placing the echoes in the resonance region [17]. This resonance effect amplifies the RCS of the drone. Compared to X-band radar, drone signals received by the L-band radar exhibit a higher signal-to-noise ratio (SNR), potentially exceeding 20 dB. This design significantly extends the detection range for drone targets. Therefore, the long-range 3D radar (L-band holographic radar GAMEKEEPER) performs well in detecting and tracking small drones, as well as providing radar tracks that aid in target classification. However, due to poorer micro-Doppler resolution within the L-band, the multi-layered drone detection system requires a secondary radar for drone recognition. Micro-Doppler spectrograms are highly frequency-sensitive, making the short-range radar (likely the X-band FMCW SQUIRE radar) a preferable choice due to its FMCW system and X-band frequency. Public papers indicate that X-band radar can not only distinguish between radar echoes from drones and birds[18] but also identify different types of drones and even differentiate between various drone payloads[19].

Despite the effectiveness of radar systems operating in the resonance and optical regions, EO sensors remain essential for identification purposes. In practical applications, recognition tasks often rely on EO/IR sensors to visually confirm radar tracks and validate drone detections through photographic evidence.

In conclusion, to enhance radar detection capabilities, improvements can be made in radar transmitters, such as enhancing antenna gain or exploring new radar bands to leverage resonance effects for amplifying drone RCS values. Additionally, advancements in signal processing algorithms, such as Track Before Detect (TBD) or Track While Scan (TWS), can enhance radar detection efficiency. For radar classification, improving micro-Doppler signal detection, optimizing track quality, and employing sensor fusion strategies are crucial for accurately identifying and classifying drone threats.

2.3 The key performance

Radar systems play a crucial role in detecting objects within their operational environment, offering autonomous operation without human intervention. Key performance metrics for radar systems include detection range, response time, and confidence factor, collectively shaping user experience and sensor effectiveness. Based on references from the FAA's reports on avian radar systems [15], existing regulations, and current radar products in the market, we propose specific requirements for radar systems aimed at enhancing detection capabilities, particularly in scenarios involving drone threats:

(1) Operate Automatically 24/7: Radar systems must operate autonomously, continuously extracting radar signals, identifying echoes, and tracking targets without human intervention. This "no man-in-the-loop" mode ensures that radar can detect targets over long ranges with minimal delay in response time. Traditionally, radar operators manually monitor radar screens, identifying potential targets amidst numerous "blips" and establishing tracks based on radar data. However, this approach is insufficient for countering drone radars, necessitating autonomous operation to ensure continuous vigilance and response capability.

(2) Sufficient Detection Range: Radar systems must provide adequate coverage to protect primary airspace and ground surfaces around critical facilities. For civil applications, especially in airport airspace protection against bird strikes, a range of approximately 5 miles is commonly recommended [20]. Similarly, the FAA designates "No Drone Zones" extending up to a 5-mile (8-km) radius from airports [21]. Military requirements may vary, with portable anti-aircraft missile systems of the U.S. Army typically limited to a radius of 6 km and altitudes from ground level to 3.5 km for brigade combat teams.

(3) Real-Time Detection Response Time (DRT): DRT is critical, encompassing the time from signal transmission to target attribute display on the radar screen. A shorter DRT indicates superior detection performance, ideally operating at the millisecond level comparable to human visual recognition times. For example, human facial recognition can occur within 120 ms, with basic classifications taking as little as 50 ms [22]. Similar to human perception, we believe that a DRT within milliseconds qualifies as real-time. Achieving this rapid DRT ensures effective real-time alert capabilities, aligning with the principle of "What You See Is What You Get" (WYSIWYG) in radar operations.

(4) Reliable Recognition Capability: Radar sensors traditionally use 1D signals for detecting, classifying, and tracking targets, which may not be intuitively interpretable by human operators. To enhance recognition of small drones, technologies like High Range Resolution Profiles (HRRP) are explored but can be impractical due to bandwidth constraints. Therefore, reliance on tracks, Doppler spectra, and micro-Doppler signals becomes crucial. However, these methods may suffer from reliability issues, especially with increased detection ranges and reduced SNR. To supplement radar data, EO/IR sensors provide visual confirmation through photographs, validating radar-detected targets. Image-based classifiers are increasingly relied upon for their ability to extract geometric information that enhances recognition reliability, crucial for effective drone

detection in diverse operational environments. Additionally, the reliability of open dataset recognition is another concern [23].

In conclusion, meeting these specified requirements ensures radar systems are equipped to effectively detect and respond to drone threats, safeguarding critical infrastructures and airspace with enhanced reliability and operational efficiency.

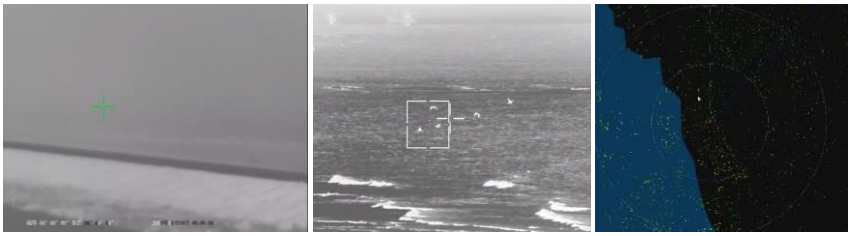
3 Case studies

3.1 Data collection

The test was conducted in Qidong, China, along the cluttered coastal environment of the Yellow Sea. The radar, mounted on a 12-m tall building, scanned horizontally over the sea. Initiated as a prototype drone detection radar project, the endeavor spanned several months in 2020, incorporating supplementary infrared sensors and an optical camera to validate Automatic Target Recognition (ATR) results. Throughout the testing period, sea conditions fluctuated between Force 3 & 5 on the Beaufort Scale. At Force 3, wind speeds ranged from 3.6-5.1 m/s, resulting in large wavelets and beginning crests with scattered whitecaps. By Force 5, wave heights escalated to 2.50-4.00 meters, with moderate waves reaching 1.2-2.4 meters, characterized by longer forms, numerous whitecaps, some spray, and wind speeds of 8.4-10.8 m/s. Various objects are within the coastal area, including ships, birds, and specifically targeted drones, detailed parameters listed in **Table 2** and corresponding images shown in the following figures.

Table 2. Targets in this test

Objects	Model	Weight (kg)	Cruising speed (m/s)	Size		Distance (~km)
				Length (m)	Width (m)	
Fixed-wing drone	Albatross11	0.3	10	0.80	1.08	11
Quad-rotor drone	Phantom 4	1.38	15	0.40	0.40	10
VTOL drone	TX25A	26	25	1.97	3.60	14
Bird	Seagull	0.5	11	0.4	1.0	9
Ship	Fishing vessel	NA	3	7	2	17



(a) **(b)** **(c)**

Fig. 3. The radar detection from signal extraction to echo identification. (a) the raw EO photo, (b) the raw IR image of the same moment, (c) the raw radar PPI image.

The radar is an X-band pulse-Doppler phased array with narrowband characteristics (**Table 3**). The radar features an active electronically scanned phased array antenna mounted on a rotating platform for full 360° azimuth coverage. Its detection response time within ten milliseconds, facilitating real-time presentation of detection and recognition results using graphic icons that indicate tracked targets. Compared to the L-band radar (Fig. 2), ours has superior detection range, DRT, and ATR capability.

Table 3. Comparison of the radars

Parameter	**Gamekeeper-16U**[9]	**WHU-X**
Frequency	L-band	X-band
Bandwidth (MHz)	~2	~12.5
Pulse Repetition Frequency (PRF)	~7.5	~5
Coherent Pulse integration (CPI) (ms)	~500	~20
Update rate (ms)	~250	~10
Max altitude (m)	900	3000
Drone detection range (km)	~5	~12
Classification	Bird, drone, aircraft, ground vehicle, surface vessel	Bird, fixed-wing drones, multi-rotor drone, VTOL drone, helicopter, ship, vehicle, pedestrian, airplane, etc.
classification using tracks	Yes	No

3.2 Method

In this section, we outline our approach to radar detection, emphasizing the critical steps of "Signal Extraction" and "Echo Recognition". Fig. 4 illustrates two signal processing chains employed in radar systems: (a) the traditional chain, and (b) our new approach integrating ATR technology.

(a) **(b)**

Fig. 4 The general signal processing chains of a radar, (a)Traditional chain [24], (b) New chain using ATR.

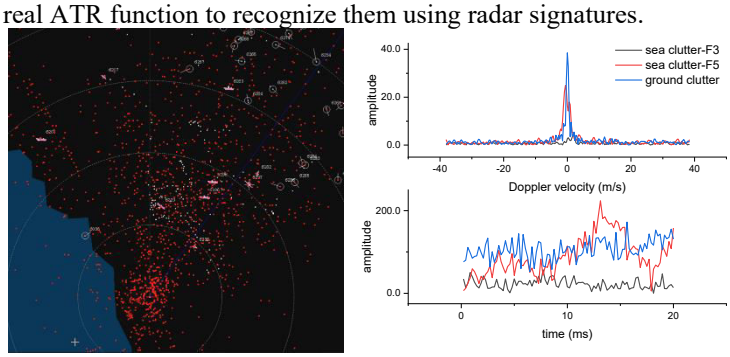
The traditional radar signal chain operates sequentially, where radar tracking precedes radar classification[24]. Radar tracks provide velocity and trajectory data, serving as signatures for object classification. This approach enhances the SNR crucial for classifying small objects like birds, drones, and wake vortices. In contrast, our novel approach employs a parallel signal processing chain that does not rely on radar tracking information upfront. Instead, the radar classifier outputs target attributes, which subsequently aid the radar tracker. We have developed an advanced radar detector utilizing the Doppler Signal-to-Clutter Ratio (DSCR) [25]. And then, we have developed an innovative radar classifier that leverages geometry mapping information derived from narrow-band radar data. This classifier effectively distinguishes echoes from a diverse array of objects such as helicopters, drones, birds, ships, vehicles, and airplanes, thereby

reducing false alarms. Specific details about our proprietary ATR classifier are protected under intellectual property policies. However, our previous projects have successfully demonstrated its capabilities in various classification tasks, including differentiating helicopters from vehicles [26], birds from drones [27], and identifying different drone types [28]. These advancements in radar detection and classification underscore our commitment to enhancing radar performance in detecting and identifying small, challenging objects in complex environments.

3.3 Typical results

Radar detection of targets always occurs in cluttered environments. Fig. 5 demonstrates one radar detection result of the area. The red "blips" in Fig. 5 are the extracted radar signals, and their number is very large. Most of them belong to background clutter. Fig. 5b shows some cases of clutter, including sea clutter and ground clutter. Different sea conditions result in varying sea clutter. "Sea clutter F3" refers to a Force 3 test, and "F5" to a Force 5 test. Thus, sea clutter F5 has higher SNR values and a wider spectrum around 0 than sea clutter F3 because Force 5 is faster and more violent than Force 3. Additionally, ground clutter has less fluctuation than sea clutter in the time series and narrower Doppler broadening. This is because the ground's natural background does not have sea waves. In brief, the first step for a radar is to extract both weak and strong signals from both targets and clutter. Otherwise, any missed signals will not be processed by the radar classifier, interrupting the signal processing chain. Most importantly, no matter how well the radar detector is designed, clutter will always be present.

After signal extraction, the most critical process is echo recognition, handled by the radar classifier. If we do not understand the meaning of the "blips" on the radar display, detection is meaningless. An area may contain many targets. Compared to the raw data in Fig. 5a, the number of targets may only be 1% of the total "blips". Generally, human operators judge tracks to classify clutter and real moving objects. Yet, many objects remain in the area. How can we choose the target of real interest? For example, we may only care about drones, yet ships and birds could still be extracted by the radar detector. In a real coastal scenario, ships and birds are common, while drones are rare. Thus, we need a real ATR function to recognize them using radar signatures.



(a) **(b)**

Fig. 5. The signal extraction from background clutter. (a) Raw "blips" on radar screen, (b) Raw radar data behind "blips".

In coastal scenarios, ships are the main surface clutter for drone detection. Compared to small drones, ships are quite large, with high RCS values. Fig. 6 shows that the ship's signal magnitude could be over 20 dB higher than the background clutter floor. Here, the radar data marked as "background" are those around the neighboring target's radar bin. To some degree, this can be seen as the similar background clutter floor. Therefore, detecting a ship, especially a fishing vessel, over a canoe is not a problem for radar. In other words, a counter-drone radar's scan must cover the surface into the low-altitude airspace. Therefore, it must classify radar echoes between surface objects, like ships or vehicles, and airborne objects, like birds and drones.

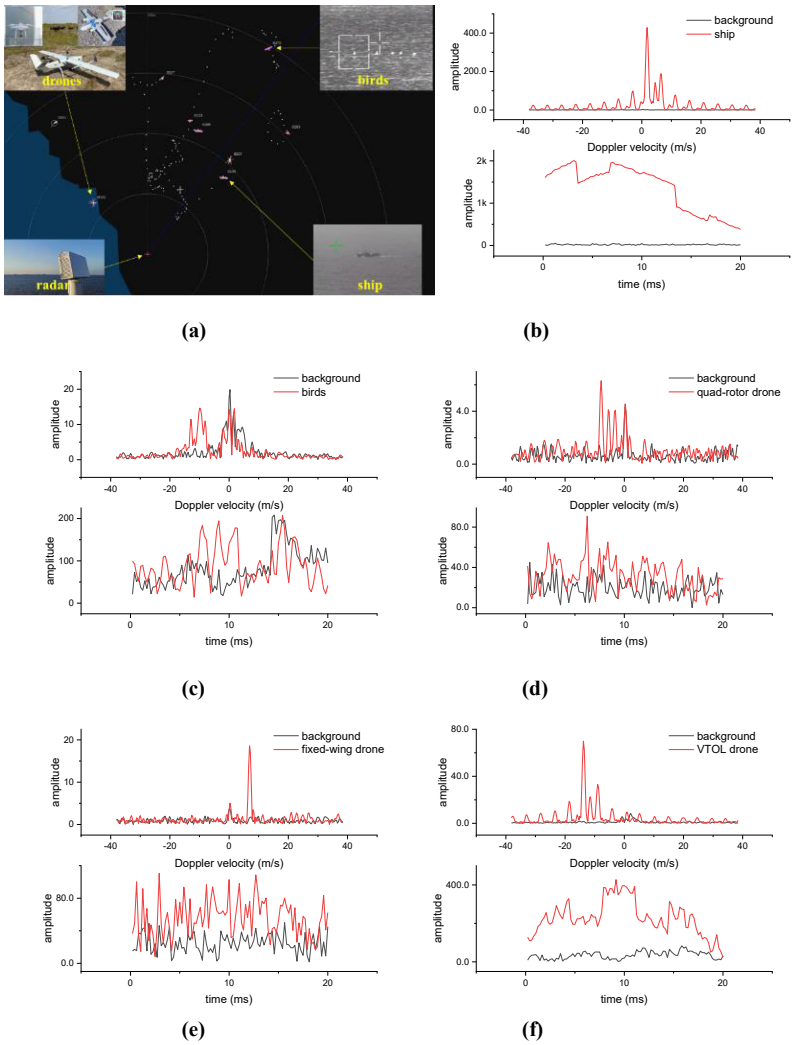


Fig. 6. Radar echo recognition between different targets, (a) recognition results with IDs, (b) ship, (c) birds, (d) quad-rotor drone,(e) fixed-wing drone, (f)VTOL drone.

Next, a counter-drone radar must distinguish radar echoes of drones from the most common airborne clutter: birds. Fig. 6 shows the detection result of a sea bird flock. In the time domain, birds have similar RCS values to small drones, which are also suppressed by the background clutter. Yet, birds can be easily detected using our DSCR detector because their flapping wings cause strong micro-Doppler in the spectrum. The birds' Doppler spectrum displays approximately four clear peaks (-12.7 m/s, -11.7 m/s, -9.9 m/s, -8.1 m/s), in addition to peaks around 0. The strongest scattering power is generated by the bird's body, observed at a Doppler velocity of -9.9 m/s, while other peaks are produced by the flapping wings. This is a new finding discussed in another paper [29]. Since birds share the same airspace with LSS drones, with similar velocities, RCS values, and even micro-Dopplers, birds are major clutter when a radar detects drones. Birds often cause false alarms for deployed radars. It is very challenging to classify the two, especially when radar dwell time is limited, yet we can classify them using some geometry information.

There are different types of drones, and a radar classifier can recognize these types. Fig. 6(d-f) shows the detection results for a quad-rotor drone, a fixed-wing drone, and a VTOL drone. Their type information is listed in **Table 2**, and their photos are shown in Fig. 6a. We have conducted more detailed research using the data collected in this area [28]. Our study demonstrates the potential of using micro-Doppler signatures modulated by different blades to improve drone detection and the identification of drone types by a drone detection radar. According to the data here, different drones have different SNR values and moving speeds due to their motion features and sizes. Essentially, a bigger drone can suppress the background clutter more effectively, making it easier to detect. Although the VTOL drone is the furthest away, it is also the largest, and its radar signals and signatures are the clearest. The most common small quad-rotor drones have signals in both the time series and spectrum that are similar to the background clutter and birds. Therefore, birds can significantly interfere with the detection of quad-rotor drones. Additionally, since the fixed-wing drone's rotating blades are perpendicular to the surface, its micro-Doppler in the spectrum and the range-Doppler images are hard to see but still present. In summary, our radar classifier can not only differentiate radar echoes of drones from birds but also distinguish between different drone types. Different types of drones have different shapes or geometries; if we identify the geometry information, we can perform the recognition.

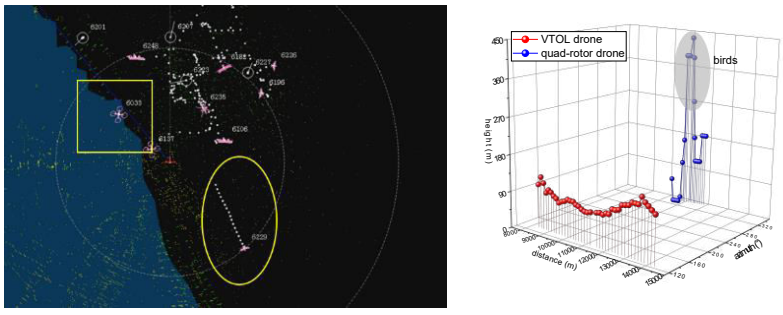
3.4 Technology Discussion

As shown in **Table 3**, our radar demonstrates a longer detection range and a faster detection response time (DRT) compared to the reference system. This improvement is attributed to advancements in our radar signal processing chain. First, we employ a DSCR detector to extract weak signals from small drones, thereby extending the detection range. Second, by eliminating the need for tracks to recognize radar echoes from small drones, our DRT is reduced to just ten seconds, significantly faster than the L-band radar's update rate. Consequently, our radar system offers quicker and more efficient drone detection.

The radar classifier has functional boundaries and cannot recognize every target as effectively as human vision. Our classifier can achieve Tier 3 recognition. The NATO

AAP-6 Glossary of Terms and Definitions, last revised in 2008, offers a comprehensive framework for understanding ATR and assessing target attributes, proposing ATR tiers for a system [30]. The six major classification steps are outlined as “Detection, Classification, Recognition, Identification, Characterization, Fingerprinting”. Currently, achieving Tier 4 remains a challenge, and we are working on projects using particularly deep learning algorithms, to investigate this topic further.

Radar tracks can easily fool the operator. Fig. 7 shows cases of radar tracks, where the dotted white lines represent the tracks. The tracking interval is about 10 seconds, and the radar tracker used the traditional TWS method. Fig. 7a demonstrates that bird tracks can appear as straight lines along the coast, easily mistaken for drones, which typically have linear flight paths. Fig. 7b illustrates a case of erroneous tracking when the radar attempted to track a small quad-rotor drone, DJI Phantom 4. Generally, this drone flies at an altitude below 200 meters. However, in some tracking frames, the "drone" suddenly appears at a high altitude of over 400 meters. The drone pilot informed the radar operator that the drone was maintaining a fixed altitude. Human observation revealed that birds above the drone were the cause of the incorrect tracking. In contrast, the large VTOL drone can be tracked more stably due to its significantly larger RCS—almost four times that of the quad-rotor drone. Regardless of the tracking algorithm's sophistication, the fundamental step is signal extraction. Therefore, high false alarm rates can lead to incorrect tracking. This indicates that traditional radar tracks can be easily fooled by clutter, especially from moving birds. Radar tracks still have value, especially when enhanced by the ATR function. In our earlier projects [31][32], we proposed the concept of Classify While Scan (CWS) technology to improve the detection performance of drone detection radar systems and enhance situational awareness.



(a) **(b)**

Fig. 7 Radar tracking cases, (a)VTOL drone, (b) Wrong tracking a small quad-rotor drone and right tracking a large VTOL drone.

4 New threats

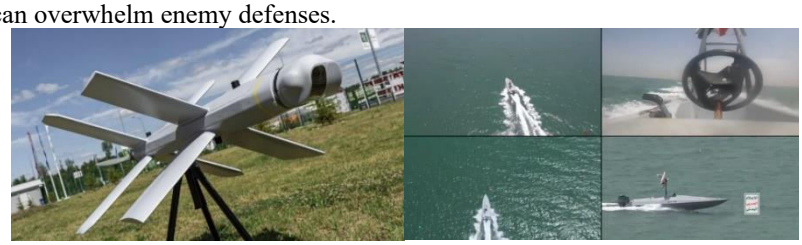
With the development of electronic components and artificial intelligence (AI), new threats are emerging, particularly from one-way attack (OWA) drones, FPV drones and sea drones. While many researchers claim that popular drone swarms pose a unique

challenge, each unit within a swarm is still just a single drone for individual radar resolution. Drone swarms are designed for saturation attacks on radar detection. However, by dividing drone swarms into separate resolution bins, they can be managed similarly to individual drones, simplifying their detection. Additionally, if a radar does not rely on track-based detection, the number of targets is not a significant issue.

We suggest that OWA drones, First-Person View (FPV) and sea drones could be a more urgent threat than drone swarms, especially considering the ongoing war crisis posed by Russia in Ukraine and the Red Sea crisis posed by Houthi rebels.

The boundary between small missiles and drones is blurring with the rise of OWA drones, or suicide drones, which have gained prominence since the 2022 Russia-Ukraine conflict (Fig. 8a). These high-speed, VTOL-capable drones revolutionize warfare with their precision and destructive power, challenging current detection systems and necessitating a review of C-UAS strategies. Additionally, FPV drones, a new type of OWA drone, play a significant role in the ongoing conflict. Ukraine's innovation with FPV drones helps counterbalance Russia's larger troop numbers. Unlike traditional multi-rotor drones, FPV drones vary in size and speed, ranging from under 1 kg to several kilograms, with speeds exceeding 100 km/h. Military-adapted FPVs might be modified for enhanced performance. This development is prompting advancements in counter-drone technology globally.

Sea drones, or Unmanned Surface Vehicles (USVs), are increasingly being developed and deployed for naval operations. Recent reports indicate that Houthi bomb boats, including some previously unseen models, are threatening Red Sea ships (Fig. 8b). These autonomous or remotely controlled vessels offer several advantages, such as reduced risk to personnel, lower operational costs, and the ability to operate in high-risk environments. USVs can be equipped with various weaponry and sensors, making them versatile tools for reconnaissance, surveillance, and direct engagement. Their ability to work in swarms further enhances their effectiveness, allowing coordinated attacks that can overwhelm enemy defenses.



(a) **(b)**

Fig. 8. Typical new drone threats for countering drone radars, (a) Russian Lancet Kamikaze OWA drone [33], (2) Houthi sea drone[34].

Another significant problem is the jamming environment and the sea background, particularly the former. The war between Russia and Ukraine has proven that jamming signals affect both drones and counter-drone solutions. Unlike radar detection, a drone can fly without communication in a jamming war condition, but radar cannot. Most current radar ATR technology for drone detection could fail when facing jamming. Small drones have a small RCS, making them hard to detect and track by radar.

Jamming signals exacerbate this issue, causing radar signatures like micro-Doppler, RCS, and speed to disappear. Therefore, anti-jamming radar technology for detecting drones is crucial. Similarly, sea clutter acts as a natural jamming environment because sea waves can absorb electromagnetism signals and disrupt the signal processing chain, resulting in no detection, classification, or tracking. In brief, countering drones or missiles in a jamming environment, whether from human jamming devices or natural backgrounds, is the real challenge for military counter-drone applications.

5 Conclusions

In this paper, we review the development of counter-drone radar systems and provide several recommendations for their improvement. We discuss three significant incidents involving hostile drone attacks: the invasion of drones at London's Gatwick Airport, the drone warfare in the Russian invasion of Ukraine, and the Red Sea crisis caused by Houthi rebels. Various counter-drone radar systems exist, each with different design philosophies, yet some perform inadequately due to the LSS characteristics of drones, posing challenges for radar detection and classification. To address these challenges, we suggest that an effective counter-drone radar must possess four key capabilities: automatic operation 24/7, sufficient detection range, real-time detection response, and reliable recognition capability. Some researchers may doubt the feasibility of such a radar. However, we present a counter-drone radar solution that meets these requirements. Coastal tests indicate that our radar can detect small drones at ranges exceeding 10 km. Additionally, since our radar classifier relies on geometric information about the targets, it ensures reliable recognition capability and maintains detection response times within milliseconds. Consequently, it operates automatically 24/7, serving as an effective sense-and-alert warning system with both WYSIWYG (What You See Is What You Get) and situational awareness functions. Looking to the future, we suggest that new drone threats include OWA drones, FPV drones and sea drones, particularly in jamming environments. Lessons learned from the ongoing war in Ukraine and the Red Sea crisis caused by Houthi rebels underscore the need for continuous advancements in counter-drone radar technology.

Acknowledgments. We appreciate the funding support from the Natural Science Foundation of Hubei Providence-Youth Program (Grant: 2023AFB130), and the 2024 Young Talent Program of the Science and Technology Think Tank of CAST (Grant: XMSB20240710063).

Disclosure of Interests. The authors have no competing interests.

References

1. Kotkova B (2022) Airport defense systems against drones attacks. In: 2022 26th Int. Conf. Circuits, Syst. Commun. Comput. IEEE, pp 85–90
2. Kunertova D (2023) The war in Ukraine shows the game-changing effect of drones depends on the game. Bull At Sci 79:95–102

3. Notteboom T, Haralambides H, Cullinane K (2024) The Red Sea Crisis: ramifications for vessel operations, shipping networks, and maritime supply chains. Marit Econ Logist 26:1–20
4. Brown AD (2023) Radar Challenges, Current Solutions, and Future Advancements for the Counter Unmanned Aerial Systems Mission. IEEE Aerosp Electron Syst Mag 38:34–50
5. Holt D, Guo W, Sun M, Panagiotakopoulos D, Warston H (2024) Deep Learning for Radar Classification. In: 2024 Int. Conf. Unmanned Aircr. Syst. pp 1208–1215
6. Sun H, Oh BS, Guo X, Lin Z (2019) Improving the Doppler Resolution of Ground-Based Surveillance Radar for Drone Detection. IEEE Trans Aerosp Electron Syst 55:3667–3673
7. Hofele F-X, Hanewinkel A (2023) Automatic Radar Target Classification A New Idea for Distinguishing Drones and Birds From the Invention to Serial Production. In: 2023 24th Int. Radar Symp. pp 1–10
8. Aievola R, Causa F, Fasano G, Manica L, Gentile G, Dubois M (2023) Conflict Detection Performance of Ground-based Radar Networks for Urban Air Mobility. In: 2023 IEEE/AIAA 42nd Digit. Avion. Syst. Conf. pp 1–9
9. White D, Jahangir M, Baker CJ, Antoniou M (2023) Urban Bird-Drone Classification with Synthetic Micro-Doppler Spectrograms. IEEE Trans Radar Syst 1–1
10. Haifawi H, Fioranelli F, Yarovoy A, Meer R van der (2023) Drone Detection & Classification with Surveillance 'Radar On-The-Move' and YOLO. In: 2023 IEEE Radar Conf. pp 1–6
11. Urzaiz FI, Gismero-Menoyo J, Asensio-López A, Quevedo ÁD de (2021) Digital Beamforming on Receive Array Calibration: Application to a Persistent X-Band Surface Surveillance Radar. IEEE Sens J 21:6752–6760
12. Mayer JE (2017) State of the Art of Airworthiness Certification. STO. NATO. int. https//www. sto. nato. int/publications/STO Meet. Proceedings/STO-MP-AVT-273/MP-AVT-273-08. pdf (accessed Jan. 5, 2020)
13. Szabolcsi R (2016) Beyond Training Minimums–A New Concept of the UAV Operator Training Program. In: Int. Conf. knowledge-based Organ. pp 560–566
14. Dempsey M (2010) Us army unmanned aircraft systems roadmap 2010-2035.
15. U.S. Department of Transportation (2005) Airport Avian Radar Systems - Advisory Circular.
16. Thales EAGLESHIELD CUAS for Airports. https://www.thalesgroup.com/en/eagleshield-cuas-airports. Accessed 22 Aug 2024
17. Melnikov VM, Lee RR, Langlieb NJ (2012) Resonance effects within S-band in echoes from birds. IEEE Geosci Remote Sens Lett 9:413–416
18. Sayed AN, Tran HH, Ramahi OM, Shaker G (2023) Radar-Based Digital Twins for Classification of UAVs and Avian Targets. In: 2023 IEEE Microwaves, Antennas, Propag. Conf. pp 1–4
19. Wit JJM De, Gusland D, Trommel RP, De Wit JJM, Gusland D, Trommel RP (2021) Radar Measurements for the Assessment of Features for Drone Characterization. In: 2020 17th Eur. Radar Conf. pp 38–41
20. (2024) Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures. U.S. Department of Transportation, Federal Aviation

Administration
21. Administration FA (2021) FAA Part 107: Small unmanned aircraft systems.
22. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans: Nineteen results all computer vision researchers should know about. Proc IEEE 94:1948–1961
23. Scheirer WJ, Rocha A de R, Sapkota A, Boult TE (2013) Toward Open Set Recognition. IEEE Trans Pattern Anal Mach Intell 35:1757–1772
24. Ahmad BI, Rogers C, Harman S, Dale H, Jahangir M, Antoniou M, Baker C, Newman M, Fioranelli F (2024) A Review of Automatic Classification of Drones Using Radar: Key Considerations, Performance Evaluation, and Prospects. IEEE Aerosp Electron Syst Mag 39:18–33
25. Gong J, Yan J, Hu H, Kong D, Li D (2023) Improved Radar Detection of Small Drones Using Doppler Signal-to-Clutter Ratio (DSCR) Detector. Drones. https://doi.org/10.3390/drones7050316
26. Gong J, Yan J, Li D (2019) Comparison of micro-Doppler signatures registered using RBM of helicopters and WSM of vehicles. IET Radar, Sonar Navig 13:1951–1955
27. Gong J, Yan J, Li D, Kong D, Hu H (2019) Interference of radar detection of drones by birds. Prog Electromagn Res M 81:1–11
28. Yan J, Hu H, Gong J, Kong D, Li D (2023) Exploring Radar Micro-Doppler Signatures for Recognition of Drone Types. Drones. https://doi.org/10.3390/drones7040280
29. Gong J, Yan J, Kong D, Chen R, Li D (2023) Formation Wing-Beat Modulation (FWM): A Tool for Quantifying Bird Flocks Using Radar Micro-Doppler Signals. arXiv Prepr. arXiv2309.15415
30. Blacknell D, Griffiths H (2013) Radar automatic target recognition (ATR) and non-cooperative target recognition (NCTR). Radar Autom Target Recognit Non-Cooperative Target Recognit. https://doi.org/10.1049/PBRA033E
31. Gong J, Li D, Yan J, Hu H, Kong D (2022) Using Classify-While-Scan (CWS) Technology to Enhance Unmanned Air Traffic Management (UTM). Drones. https://doi.org/10.3390/drones6090224
32. Gong J, Li D, Yan J, Hu H, Kong D (2023) Avian radar system using phased array radar technologies. In: Proc. IEEE Radar Conf. pp 1–6
33. Hambling D Russia Steps Up Deployment Of Lancet Kamikaze Drones, But How Effective Are They? https://www.forbes.com/sites/davidhambling/2024/06/25/russia-steps-up-deployment-of-lancet-kamikaze-drones-but-how-effective-are-they/. Accessed 22 Aug 2024
34. Epstein J Houthi bomb boats, including some not seen before, are threatening Red Sea ships while the US Navy's aircraft carriers are away. https://www.businessinsider.com/houthi-drone-boats-trouble-red-sea-us-aircraft-carriers-gone-2024-7.

Development of a classification model for UAVs and birds based on the YOLOv9 neural network to improve Anti-drone systems

Vladislav Semenyuk[1][orcid.org/0000-0002-8580-7326], Ildar Kurmashev[1][0000-0001-9872-7483], Dmitriy Alyoshin[1][0000-0002-5985-0523], Liliya Kurmasheva[1][0000-0001-5392-5873], Alessandro Cantelli-Forti[2][0000-0002-6943-2632]

[1] M. Kozybayev North Kazakhstan University, Pushkin street, 86, Petropavlovsk, 150000, Kazakhstan
[2] RaSS (Radar and Surveillance Systems) National Laboratory, 56124 - Pisa, Italy

Abstract. The article presents the materials of the development of a model for classification and recognition of UAVs and birds based on the neural network of the YOLOv9 architecture in the optoelectronic channels of Anti-drone systems. To train the neural network, a dataset was prepared in the form of annotated images of UAVs and birds. The total number, taking into account augmentation, was 5265 images. The authors implemented training, verification and testing of neural networks in the Windows 11 operating system, in the Python 3.10.8 runtime environment and the Pycharm 2024 development environment. The training process was carried out on the basis of the AD103 graphics processor of the NVIDIA GeForce RTX 4080 video card with support for CUDA Toolkit 12.1. As a result of training the neural network, the following metrics were obtained: mAP50-95: 0.59; mAP50: 0.95; Recall: 0.89; Precision: 0.95. According to these indicators, the trained model outperforms the UAV and bird recognition and classification models trained on the basis of YOLOv2, YOLOv4, YOLOv5, YOLOv7 and YOLOX. The inference results on two videos with DJI Inspire 2 and DJI Mini 3 UAV flights showed FPS values of 131 and 119, respectively. It was found that, due to the obtained accuracy and FPS metrics, the trained YOLOv9 model can be used as a module for recognizing and classifying UAVs and birds in real time in the optoelectronic surveillance channels of Anti-drone systems.

Keywords: Anti-drone, Sensor fusion, Deep learning, Drones, YOLO, Neural networks.

1 Introduction

Currently, Anti-drone systems are actively used to solve problems of detection, classification and neutralization of UAVs (drones). Such a need is due to the growing number of incidents of using these devices for criminal purposes. Examples include violation of airport airspace, espionage, mass attacks on critical facilities for military purposes,

delivery of prohibited items, and organization of failures in security systems. In this regard, the development of new methods for combating UAVs and the improvement of existing technologies for detection, classification and elimination of UAVs is relevant. As a result of the literature review of the Scopus and Web of Science databases, optoelectronic, acoustic, radio frequency, radar and combined (Sensor Fusion) methods are used to detect and classify UAVs, represented by the corresponding software and hardware solutions. Optoelectronic systems use cameras [1], laser sensors [2], thermal imagers [3] to accurately detect and track UAVs, but their effectiveness may be reduced in low visibility and lighting conditions. Radar systems [4, 5], on the other hand, operate based on radio waves and are capable of detecting objects in all weather conditions and at any time of day, although their accuracy in recognizing small objects may be limited. Radio frequency systems (RF based) [6, 7] detect UAV control and communication signals, which allows them to be detected even in the absence of visual contact, but they rely on the presence of radio signals and may encounter interference. Acoustic systems [8, 9] use microphones to detect sounds emitted by UAVs, which makes them useful in low visibility conditions, but their range and sensitivity may be limited by environmental noise. To improve detection accuracy and reliability, sensor fusion technology [10–12] is often used, which combines the data stream from different sensors. This allows for higher accuracy and reliability, although integration and data processing may be complex. This technology is implemented in modern Anti-drone systems, such as Elbit Systems ReDrone [13], DedroneRapidResponse [14] and others.

Accurate recognition and classification of UAVs relative to other objects is provided by the software component of the Anti-drone systems - artificial intelligence, which is represented by machine learning (ML) and deep leaning (DL) algorithms. In the optoelectronic channels of the Sensor Fusion systems, which are considered indispensable due to the accuracy of providing visual data, computer vision algorithms of the YOLO architecture are implemented. This algorithm, along with Faster R-CNN, SSD, RetinaNet and EfficientDet, is used as a visual detector of objects in real time. Due to key features such as single-stage processing, dividing images into a grid, joint prediction of different classes, high accuracy and speed, YOLO is more effective in solving problems of recognition and classification of objects, including UAVs. The authors [15–18] used various YOLO models to train neural networks on user datasets in the form of images of UAVs of different types, birds, etc. In [15], the authors trained the YOLOv4 model to recognize UAVs and birds, achieving the following average accuracy rates: mAP50 – 74.36%; precision – 0.95; Recall – 0.68; F1 – 0.79.

When tested on videos of two types of UAVs, DJI Phantom III and DJI Mavic Pro, the trained model achieved 20.5 and 19 FPS (frames per second), respectively, on inference. In [16], the authors used an earlier YOLOv2 model and achieved an mAP50 of 74.97%. The YOLOv5 model from [17] outperformed the previous model [16] by 15.4%. Higher-performance models of the YOLO architecture, such as YOLOX, YOLOv7, YOLOv8, are studied in [18]. YOLOv8 is a more advanced version of the previous models, thanks to new features and improvements implemented by the developers of Ultralytics. The new backbone network, anchorless detection head, and loss function contributed to high-quality training of the model with an mAP50 of 95.3%. The accuracy of UAV recognition and classification, which are characterized by the

mAP50, mAP50-95, Precision and Recall metrics, are limited by the loss of information in successive layers of deep neural networks. This problem can be solved by implementing programmable gradient information (PGI) and the architecture of efficient layer aggregation network (GELAN).

In order to improve the model for recognizing and classifying UAVs in optoelectronic detection channels of Anti-drone systems by increasing the accuracy indicators, the following objectives must be completed within the framework of this study:

– Prepare a dataset of UAV and bird images for training the experimental YOLOv9 neural network model;
– Train the YOLOv9 neural network model to determine the accuracy indicators;
– Test the model on inference to determine FPS.

The results obtained allow us to draw a conclusion about the effectiveness of using the YOLOv9 neural network model for recognizing and classifying UAVs and birds.

2 Research methods

The training of the neural network model for UAV and bird classification will be based on the pre-trained YOLOv9 algorithm. This algorithm overcomes the shortcomings of methods for overcoming information loss, such as reversible architectures, masking modeling, and the concept of deep supervision, by implementing PGI. PGI (see **Ошибка! Источник ссылки не найден.**) is based on gradient generation using an auxiliary reversible branch, which allows avoiding loss at semantic levels without additional computational costs.

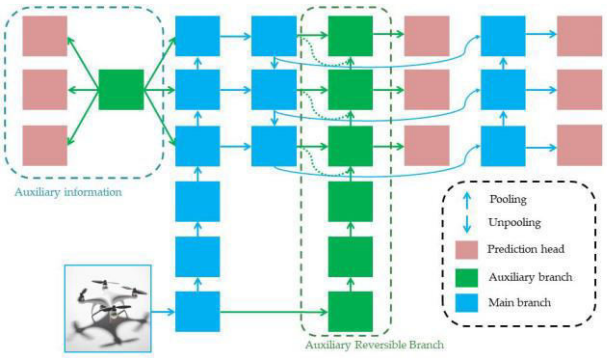


Fig. 1. PGI Architecture diagram.

The main structural components of PGI, in addition to the above-mentioned auxiliary reversible branch, are the main branch and multi-level auxiliary information. The main branch is used to organize logical inference, and multi-level auxiliary information solves the problems of error accumulation due to deep observation, which is essential for training lightweight models of the YOLOv9 architecture. The presence of GELAN in the YOLOv9 neural network ensures the creation of multi-scale feature maps for

class prediction, optimization of parameters, computational complexity, accuracy and inference speed. This advantage is due to the combination of two neural network architectures CSPNet and ELAN, which provide gradient path planning. This solution allows users to update computing units for any logical output devices without significant performance losses.

To train the YOLOv9 model on custom datasets, the following steps must be completed:

- Prepare a dataset of images for two classes (UAVs and birds);
- Annotate classification objects in each image;
- Data augmentation;
- Distribute the dataset into three parts for training, validation, and testing;
- Export data for training in a special design environment;
- Select hyperparameters and start training.

To successfully train the neural network model, it is necessary to prepare a dataset for two classes of objects that will meet the criteria of data diversity (a variety of UAV and bird images representing different lighting conditions, weather conditions, angles and backgrounds), class balance, high image quality, image normalization to one size (e.g. 416 x 416; 640 x 640). Images for a custom dataset can be found in open sources such as Roboflow, Kaggle, Ultralytics, GitHub, and also use custom data (photos, videos) of UAV and bird flights. Annotation is a markup of images with special bounded boxes. This frame is defined by the coordinates of the upper left corner and lower right corner or the center coordinates and dimensions (width and height). For each image, an annotation file is created that contains information about the objects in the image. The format of the annotation file can be different, but usually YOLO uses a text format, where the class of the object and the coordinates of the bounding box are indicated for each object. For YOLOv9, information on annotated objects is stored in a separate ".txt" file. Roboflow.com is used as a service to load, store, process and annotate class objects, as shown in Figure 2. Augmentation technology is used to increase the variability and quantity of data.

This technology is a technique for artificially increasing the size and diversity of a data set by applying various transformations to the original data. This technique is widely used in machine learning, especially in computer vision, to improve the quality of models and prevent overfitting.

In the case of images, augmentation involves applying various transformations. Geometrical transformations include rotating an image by a random angle, changing its scale, shifting it horizontally or vertically, flipping it horizontally or vertically, as well as bending and distorting it. Color transformations involve adjusting the brightness, changing the contrast, saturation, and color hues of an image. Also, various types of noise can be added to an image, blurring can be applied, or the image can be randomly cropped and resized back to its original proportions.

(a) UAVs (b) Birds

Fig. 2. Annotating objects in Roboflow

In addition to these methods, there are other techniques such as excluding random square areas from an image (cutout) and mixing two images and their labels to create a new sample (mixup). Using data augmentation helps models become more robust to various types of distortions, which ultimately improves their ability to generalize to new data and leads to better performance on real data. Using this technology and the Roboflow interface, the Grayscale and Blur augmentation methods were applied. In addition to training, the model undergoes validation and testing stages, in connection with which the data set is distributed in percentage terms of 82/13/6 (training, validation and testing, respectively). A special annotation format YOLOv9 was selected to export the dataset to the neural network training project.

The training, verification, and testing of neural networks were implemented in the Windows 11 operating system, in the Python 3.10.8 runtime, and the Pycharm 2024 development environment. The training process was carried out on the AD103 GPU of the NVIDIA GeForce RTX 4080 video card with support for CUDA Toolkit 12.1. The program code was written and edited using the Ultralytics YOLOv8 framework, which contains YOLOv9 neural network models pre-trained on the COCO dataset. Among the five YOLOv9 models (YOLOv9t, YOLOv9s, YOLOv9m, YOLOv9c, YOLOv9e), YOLOv9c was selected for the experiment. This model was pre-trained on the COCO dataset with the following parameters: mAP50-95 - 0.53; mAP50 - 0.702; params (number of parameters) – 25.5 m (millions); FLOPs – 102.8. The following hyperparameters are set to train the neural network on the custom dataset: number of epochs: 100; batch size: 16; learning rate: 0.001; momentum: 0.9; weight decay: 0.0005 and image size: 640. The trained YOLOv9c model has a “best.pt” file size of 88 MB. The neural network was tested on inference using two test videos with DJI Inspire 2 and DJI Mini 2 UAV flights.

3 Results

3.1 Results of preparing a dataset for training the YOLOv9c neural network

Fig 3 shows the interface of the prepared dataset in the Roboflow.com service.

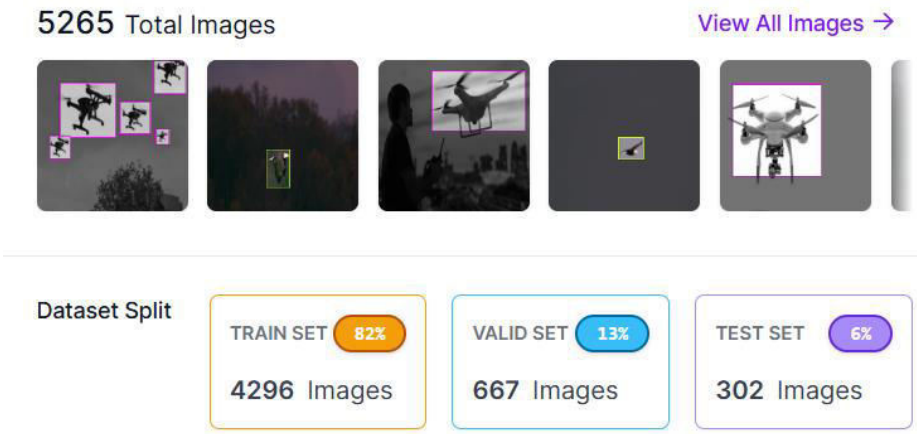


Fig. 3. UAV and bird dataset prepared by Roboflow.com.

The prepared dataset was used to train the neural network of the YOLOv9c architecture.

3.2 Results of training the YOLOv9c neural network on a custom dataset

Fig 4 a-d shows the metrics of the training results of the YOLOv9c neural network model.

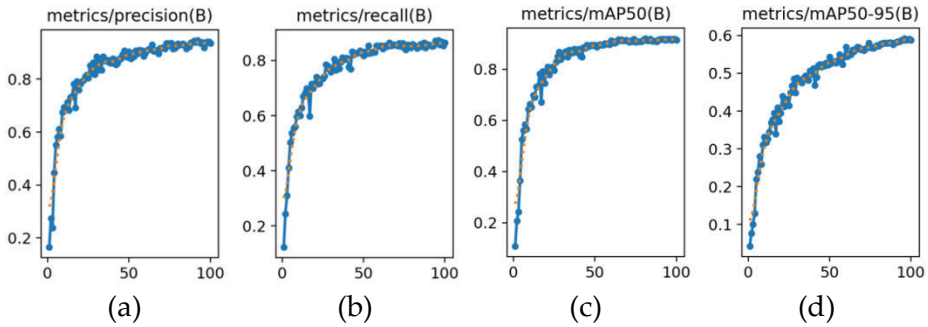


Fig. 4. Metrics of the results of training the YOLOv9c neural network for 100 epochs (Ox-axis): a – Precision; b – Recall; c – mAP50; d – mAP50-95.

The trained model was tested on inference using two videos to determine the average Latency (or FPS) value.

3.3 Inference test results

Fig 5 a,b show frames of inference of the trained YOLOv9c neural network model on two videos with the flight of DJI Inspire 2 and DJI Mini 3, respectively.

(a) (b)

Fig. 5. Frames from the inference of the trained YOLOv9c model: a) frame from the video of the DJI Inspire 2 flight; b) frame from the flight of the DJI Mini 3.

The inference results are used to estimate the average FPS presented in Section 4 Discussion.

4 Discussion

To train the YOLOv9c neural network, a dataset of 5265 annotated images (Fig 3) was prepared in the Roboflow.com service. Of this set, 4296 images (82%) are intended for training, 667 images (13%) for validation, 302 files (6%) for testing.

As a result of training, the following maximum metric values were obtained (see Fig. 4 a-d):

- mAP50-95: 0.59;
- mAP50: 0.95;
- Recall: 0.89;
- Precision: 0.95.

The obtained mAP50 values exceed the accuracy of the trained model from [15, 16] by 20%; [17] by 4.6%. The YOLOv9s model corresponds to the trained YOLOv8 neural network presented in [18] in this indicator. Thus, the YOLOv9 architecture neural network models, along with YOLOv8, are able to improve the accuracy of recognition and classification of UAVs and birds through optoelectronic surveillance cameras. As a result of inference testing, the average Latency value for the first video was 7.6 ms and 8.4 ms for the second video. These indicators depend on the GPU characteristics and for less productive computing devices, it is recommended to use lighter versions

such as YOLOv9t, YOLOv9s and YOLOv9m. The obtained Latency values prove the efficiency of using YOLOv9 as a basic software component for recognizing and classifying UAVs in optoelectronic surveillance channels of Anti-drone systems. In the future, it is planned to test the efficiency of the developed model in the tasks of segmentation of UAV and bird tracking in real time. Such an approach will not only classify UAVs from other objects, but also potentially allow them to be precisely neutralized using automated electronic countermeasure systems. To confirm the efficiency of using the developed model in Sensor Fusion systems, it is planned to develop an experimental model of a combined UAV and bird detection and classification system by integrating an optical-electronic video surveillance channel, a FMCW-radar, acoustic and RF-based sensors.

5 Conclusion

1) As part of the study of the possibility of using the YOLOv9 neural network to recognize and classify UAVs and birds using the Roboflow.com service, a dataset of 5265 annotated images was prepared.

2) Based on the AD103 graphics processor of the NVIDIA GeForce RTX 4080 video card with support for CUDA Toolkit 12.1, the YOLOv9c neural network was trained and metrics were obtained demonstrating a relatively high accuracy of UAV and bird classification in comparison with previous models of the YOLO architecture.

3) As a result of testing the trained YOLOv9c neural network on inference, average FPS values of 131 and 119 were obtained, which, along with high accuracy, proves the possibility of using this model as a module for recognizing and classifying UAVs and birds in real time in the optoelectronic surveillance channels of Anti-drone systems.

4) The results of this study will be useful to developers of Anti-drone systems.

Acknowledgments. This research is funded by the Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AR19679009).

Disclosure of Interests. The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

References

1. Mahdavi F., Rajabi R. (2020). Drone Detection Using Convolutional Neural Networks, 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), Mashhad, Iran, 2020, pp. 1-5, DOI: 10.1109/ICSPIS51611.2020.9349620.
2. Hammer M., Borgmann B., Hebel M., Arens M. (2020). Image-based classification of small flying objects detected in LiDAR point clouds. Proceedings of SPIE - The International Society for Optical Engineering, 11410, 1141002. DOI: 10.1117/12.2557246.
3. Mebtouche N.E.-D., Baha N. (2022). Robust UAV detection based on saliency cues and magnified features on thermal images. Multimedia Tools and Applications. 82(13), pp. 20039-20058. DOI: 10.1007/s11042-022-14271-3.

4. Beasley P., Ritchie M., Griffiths H., Miceli W., Inggs M., Lewis S., Kahn B. (2020). Multi-static Radar Measurements of Uavs at X-Band and L-Band. IEEE National Radar Conference - Proceedings, 2020-September, art. no. 9266444. DOI: 10.1109/RadarConf2043947.2020.9266444.
5. Teo M.I., Seow C.K., Wen K. (2021). 5G Radar and Wi-Fi Based Machine Learning on Drone Detection and Localization. 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, China, 2021, pp. 875-880, DOI: 10.1109/ICCCS52626.2021.9449224.
6. Flak P., Czyba R. (2023). RF Drone Detection System Based on a Distributed Sensor Grid With Remote Hardware-Accelerated Signal Processing. IEEE Access. PP. 1-1. DOI: 10.1109/ACCESS.2023.3340133.
7. Yang S., Luo Y., Miao W., Ge C., Sun W., Luo C. (2021). RF Signal-Based UAV Detection and Mode Classification: A Joint Feature Engineering Generator and Multi-Channel Deep Neural Network Approach. Entropy. 23. 1678. DOI: 10.3390/e23121678.
8. Al-Emadi S., Al-Ali A., Al-Ali A. (2021) Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks. Sensors 2021, 21, 4953. DOI: 10.3390/s21154953.
9. Salman S., Mir J., Farooq M.T., Malik A.N., Haleemdeen R. (2021) Machine learning inspired efficient audio drone detection using acoustic features. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 335–339. DOI: 10.1109/IBCAST51254.2021.9393232.
10. Svanström F., Alonso-Fernandez F., Englund C. (2021). A Dataset for Multi-Sensor Drone Detection. Data in Brief. 39. DOI: 10.1016/j.dib.2021.107521.
11. Semenyuk V., Kurmashev I., Lupidi A., Cantelli-Forti A. (2023). Developing the GoogleNet neural network for the detection and recognition of unmanned aerial vehicles in the Data Fusion System. Eastern-European Journal of Enterprise Technologies, 2(9-122), pp. 16–25. DOI: 10.15587/1729-4061.2023.276175.
12. Jajaga E., Rushiti V., Ramadani B., Pavleski D., Cantelli-Forti A., Stojkovska B., Petrovska O. (2022). An Image-Based Classification Module for Data Fusion Anti-drone System. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13374 LNCS, pp. 422–433 DOI: 10.1007/978-3-031-13324-4_36.
13. Elbit Systems (2024). ReDrone. [online] Elbit Systems. Available at: https://elbit-systems.com/product/redrone/ [Accessed 22 June, 2024].
14. Dedrone (2024). DedroneRapidResponse: Multi-Layered Mobile Drone Detection Unit. [online] Dedrone. Available at: https://www.dedrone.com/solutions/dedrone-rapid-response [Accessed 22 June, 2024].
15. Singha S., Aydin B. (2021). Automated Drone Detection Using YOLO v4. Drones. September 2021. DOI: 10.3390/drones5030095.
16. Seidaliyeva U., Alduraibi M., Ilipbayeva L., Almagambetov A. (2020). Detection of Loaded and Unloaded UAV Using Deep Neural Network. In Proceedings of the 2020 Fourth IEEE International Conference on Ro-botic Computing (IRC), Taichung, Taiwan, 9–11 November 2020; pp. 490–494. DOI: 10.1109/IRC.2020.00093.
17. Aydin B., Singha S. (2023). Drone Detection Using YOLO v5. Eng. 4. DOI: 10.3390/eng4010025.
18. Zhai X., Huang Z., Li T., Liu H., Wang S. (2023). YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. Electronics. 12. 3664. DOI: 10.3390/electronics12173664.

A Novel Data Fusion Algorithm to Improve the Detection and Tracking of "Killer" Drones in Urban Environment

Bhaskar Ahuja[1,2*], Walter Matta[3], Ajeet Kumar[2], Alessandro Cantelli-Forti[2]

[1]University of Trento, Italy.
[2]Radar and Surveillance Systems National Laboratory, CNIT, Pisa, 56124, Italy.
[3]Link Campus University, Rome, Italy.

*Corresponding author(s). E-mail(s): bhaskar.ahuja@unitn.it;
Contributing authors: w.matta@unilink.it; ajeet.kumar@cnit.it;
alessandro.cantelli.forti@cnit.it;

Abstract

A data fusion algorithm that incorporates joined probability between observed raw data from multiple sensors is described. Remote sensors of various kinds transmit position information to a central station where data are cinematically fused to create a composite measure. It follows that the problem is to find spatial coordinates of that point where the likelihood of finding a target is maximum in that instant. If each raw data has its own associated probability, it's possible to merge all of the sensor's likelihood to obtain the joined probability. Because raw sensor measurements can be considered independent, the target spatial coordinates with maximum likelihood are the most frequent value of joined probability distribution. This method significantly enhances tracking accuracy, providing superior target position estimates compared to single-sensor approaches. This methodology is particularly relevant for anti-drone systems in urban scenarios aimed at protecting critical infrastructures. A method to estimate the standard deviation of each sensor's associated probability is also included.

Keywords: Multi-sensor fusion, Data fusion, Drone detection, Killer drone

1 Introduction

The data fusion problem has attracted much attention because of its potential applicability in traffic management, combat identification system design, airborne surveillance, counter drones, missile defense and so on [1]. Also, remote sensing systems often involve observing phenomena of interest using multiple sensors of various types and located at different spatial positions [2]. Significant strides have been achieved by using multiple sensors such as radars and electro-optics (EO) for tracking debris and active satellites [3, 4]. With the integration of artificial intelligence in data fusion, advances in main application fields such as remote sensing, including object identification, classification, change detection, and maneuvering target tracking, are described in [5]. Additionally, deep learning algorithms have been proven useful in real-time tracking of Unmanned Aerial Vehicle (UAV) by optimising the YOLO4 (You Only Look Once V4) target detector algorithm [6]. The following authors have shown the latest progress on the image-based fusion methods for killer drone detection [7–9]. Furthermore, the literature highlights significant advancements in imaging 3D structures using Inverse Synthetic Aperture Radar (ISAR) combined with multiple sensor data fusion techniques, which enhance the accuracy and resolution of remote sensing applications in complex environments [10–12]. Despite numerous architectures for sensor fusion, these applications necessitate integrating raw data from multiple dissimilar sensors situated at remote locations [13]. The information can then be transmitted via communication links to a central station. At the central station, raw sensor data can be fused to produce a composite value of superior quality, which is not obtainable from either a single sensor or a single platform. Consequently, this methodology facilitates a more robust and reliable estimation process, leading to more accurate and insightful conclusions. This paper refers to anti-drone systems for urban areas scenarios and, in general, aimed at protecting critical infrastructures. These systems must have a high degree of flexibility and potential for evolution, which allows them to manage the progressive increase in the kinematic and offense capabilities of “killer” drones. In particular, this anti-drone system consists of a network of mini-radars (LPI polarimetric with FMCW or noise-like waveform, newly designed, with low environmental impact, on-demand imaging capabilities, fully digital, easily reconfigurable at a high level of flexibility and operating regardless of weather conditions) to be appropriately positioned on the ground in the area to be protected. The low cost of each element makes it possible to develop a network that covers practically the entire urban area at the city or neighborhood level, ensuring an almost uniform detection and tracking capacity for each “killer” drone (i.e., target) to be protected. To improve the neutralization capacity of the killer drone, it is absolutely important to improve the tracking through advanced data fusion algorithms capable of ”fuse” the data incoming of the different radars of the network to generate the ”optimal trajectory of the target. Therefore, this paper is concerned with an algorithm for data association based on joint probability. The key idea is to find the spatial coordinates of the point where the likelihood of finding a target is maximum. By merging the probability distribution associated with each sensor’s position value, obtaining a more accurate target position value is possible than a single sensor position information. The algorithm starts by estimating

the standard deviation of the probability associated with each sensor [14]. This superior quality value also depends on the estimated standard deviation of a probability distribution. In the following, the joined probability algorithm results are compared with other data fusion techniques [15–18]. The paper is structured as follows: Section 2 describes the preliminary knowledge required and the mathematical background, Section 3 details the numerical estimation techniques employed, and Section 4 presents the results. Finally, in Section 5, we summarize the relevant conclusions and propose avenues for future research.

2 Preliminary

If we consider n measurement sequence of a same quantity x denoted by $x_1, ..., x_n$, the arithmetic average also known as mean (μ_x) of that measured quantity is

$$\mu_x = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

The higher the measurement number n, the nearer the arithmetic average is to the exact value of that quantity (or to the most probable value of that quantity). The difference $(x - x_i)$ is called deviation and can be positive or negative. It comes from experience that little deviations are more frequent than bigger deviations; more exactly, if the absolute value of deviations increases, their frequency decreases, and so the probability that they happen decreases. If we trace the probability density function of deviations we obtain a Gaussian curve that is the narrower the more accurate is the method and instrumentation utilized [19]. A great significance is also the definition of standard deviation, which is shown here in eqn (2)

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} (x - \mu_x)^2}{n} \tag{2}$$

The error theory says that if the stochastic errors prevailed over deterministic ones, the measure of a quantity obtained by infinite measurement is written in the form of eqn (3)

$$x = \mu_x \pm \sigma_x \tag{3}$$

and the associated probability is given by

$$P(\mu_x - \sigma_x \leq x \leq \mu_x + \sigma_x) = \int_{\mu_x - \sigma_x}^{\mu_x + \sigma_x} f_{gauss}(x)dx \tag{4}$$

where f_{gauss} is the probability Gauss function.

$$f_{gauss} \equiv function(x, \mu_x, \sigma_x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}} \tag{5}$$

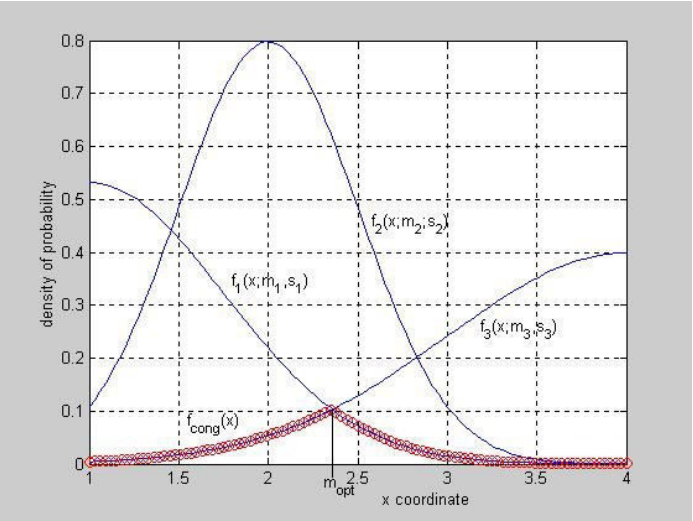


Fig. 1. Probability Gaussian distribution for three sensors

2.1 Mathematical Model

If sensor coordinates are known, the problem is to find spatial coordinates of that point where the likelihood of finding a target is maximum in that instant. Each sensor will give the target coordinates in each temporal instant. The inputs of the model are: Cartesian coordinates triad from each sensor $(x^s(t), y^s(t), z^s(t))$, Gaussian error triad, in Cartesian coordinates associated to each sensor triad $(e_x^s(t), e_y^s(t), e_z^s(t))$, where s represent sensor identification number.

As mentioned before, every coordinate value given by each sensor is to be meant as the average value of infinite measurements that are repeated with the same boundary conditions and given in eqn (6)

$$\begin{cases} x^s(t) \equiv \mu_x(t) \\ y^s(t) \equiv \mu_y(t) \quad s = (1, 2, , , , S) \\ z^s(t) \equiv \mu_z(t) \end{cases} \tag{6}$$

At the same time sensor errors (e_x, e_y, e_z) are known. It means that the error value of each sensor coordinates is the sensor's mean quadratic divergence in that instant and it is shown in eqn (7)

$$\begin{cases} \sigma_x^s(t) \equiv e_x(t) \\ \sigma_y^s(t) \equiv e_y(t) \quad s = (1, 2, , , , S) \\ \sigma_z^s(t) \equiv e_z(t) \end{cases} \tag{7}$$

It follows that for every temporal instant, we have s probability Gaussian distributions for x axis. Similar distribution curves can be written for y and z axes.

$$f(x,t) = \frac{1}{\sqrt{2\pi(\sigma_x^s(t))^2}} e^{\frac{-(x(t)-\mu_x^s(t))^2}{2(\sigma_x^s(t))^2}} \quad s = (1,2,,,,S) \tag{8}$$

Since the s sensor measurements can be considered independent of each other, the target coordinates value is the most frequently joined probability value as we can see from Figure 1, the best target position at t time along x coordinate is $x(t) = x^{opt}(t)$ with probability $f_{cong} = f(x^{opt})$.

The joined probability is the intersection area of the sensor's Gaussian measurement distributions, the joined probability value is the maximum of that area along y axes, and the most frequent value is the coordinate value of the maximum along x axes. The subsequent section will explain the numerical algorithms used.

3 Numerical Algorithm

We consider only the x coordinate here for simplicity. As described in Section 2.1, for every temporal instant t, the mean value and the variance of each Gaussian distribution (μ_x, σ_x) are given, and simultaneously, the most frequent value x is contained in the interval.

$$x \in [min(\mu_1, \mu_2, .., \mu_s), max(\mu_1, \mu_2, .., \mu_s)] \tag{9}$$

To minimize the computational load, we can determine the optimal or most frequent value by searching the zero of the joined probability function's first derivative. As shown in Figure 2, the joined probability function has a discontinuity in correspondence with the optimal value, so the interval bisection method can be applied [20]. The numerical steps are described in detail below:

- Determine the interval (prime interval) which contains the zero of the joined probability function.

$$\begin{aligned} a^{(0)} &= \min_{s=1,2,,,S} \{\mu_s\} \\ b^{(0)} &= \max_{s=1,2,,,S} \{\mu_s\} \end{aligned} \tag{10}$$

- Calculate the first derivative function $f_{cong}(x)$ on interval extremes.

$$\begin{aligned} i &= \min_{s=1,2,,,S} \left\{ f_s(a^{(0)} \right\} \\ f'_{cong}(a^{(0)}) &= \left(\frac{df_i}{dx} \right)_{x=a^{(0)}} = -\frac{a^{(0)} - \mu_i}{\sigma_i^2} . f_i(a^{(0)}) \end{aligned} \tag{11}$$

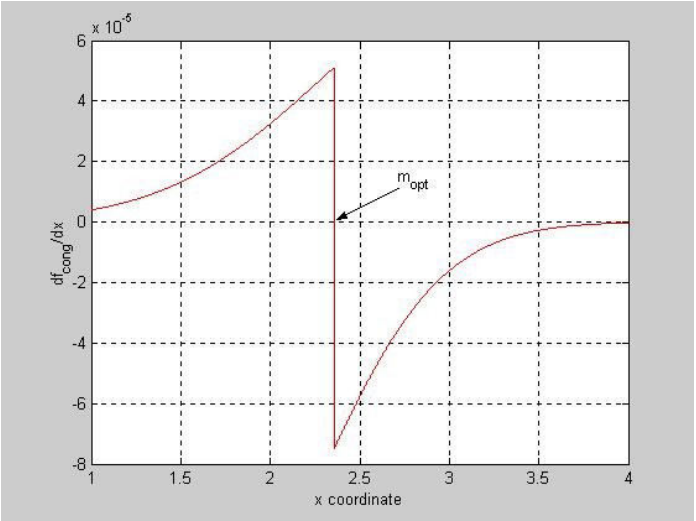


Fig. 2. Plot of the first derivative of function shown only for x direction

$$j = \min_{s=1,2,,,S} \left\{ f_s(b^{(0)}) \right\}$$
$$f'_{cong}(b^{(0)}) = \left(\frac{df_j}{dx} \right)_{x=b^{(0)}} = -\frac{b^{(0)} - \mu_j}{\sigma_i^2} . f_j(b^{(0)}) \tag{12}$$

- Calculate zero at first iteration by prime interval bisection.

$$z^{(1)} = a^{(0)} + \frac{b^{(0)} - a^{(0)}}{2} \tag{13}$$

- Determine the first iteration interval where zero is contained.

$$h = \min_{s=1,2,,S} \left\{ f_s(z^{(1)}) \right\}$$
$$f'_{cong}(z^{(1)}) = \left(\frac{df_h}{dx} \right)_{x=z^{(1)}} = -\frac{z^{(1)} - \mu_h}{\sigma_h^2} . f_h(z^{(1)}) \tag{14}$$

Then, the first derivative signs in interval extremes and middle abscissa will be compared to determine the first iteration interval by applying these conditions.

if $f'_{cong}(z^{(1)}) == f'_{cong}(b^{(0)})$
{
$b^{(1)} = z^{(1)};$
$a^{(1)} = a^{(0)};$
}

else
{
$a^{(1)} = z^{(1)}$;
$b^{(1)} = b^{(0)}$;
}
end

• Calculate zero at the second iteration as

$$z^{(2)} = a^{(1)} + \frac{b^{(1)} - a^{(1)}}{2} \tag{15}$$

• Iterate the procedure until the following criteria are satisfied

$$|z^{k+1} - z^k| < \epsilon \tag{16}$$

where ϵ is a pre-established allowance.

4 Results

The simulation of four sensors is considered for validation of the proposed algorithm. Each sensor has an arbitrary error from the true trajectory and has its detection capbility. These sensors are ground-based phase array radars which can be electronically rotated in azimuth- and elevation-direction. The field of view of radars is kept to be 5 degrees. The first step is to generate sensor detections from a multi-radar network.

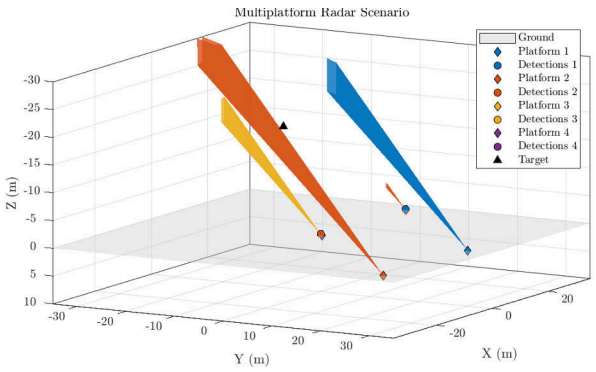


Fig. 3. Snapshot of simulation with 4 radars and 1 target

After generating these synthetic measurements with random Gaussian noise, such as range, azimuth, and elevations, the position of the target can be calculated in the sensor reference frame. However, if we know our sensors' location perfectly, we can convert the measurement to a common reference frame. A snapshot of the four sensors and a target simulation has been shown in Figure 3. The second step is to generate the target trajectory as seen by each radar and generate the probability distribution

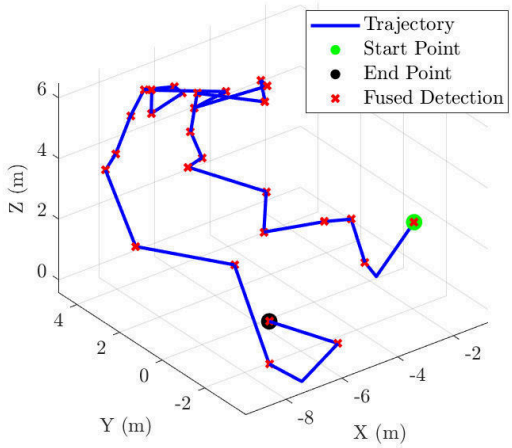


Fig. 4. Target trajectory plot with fused detection points

curve similar to the Figure 1 for each point of time. Because each radar has its field of view, we may miss some data from a sensor if there is a high-speed target and it goes out of the field of view. The last and final step is to run the algorithm discussed in Section 3 and fuse the data with joint probability approach. Fused data points will be those where likelihood of finding the target is maximum. Target's true trajectory with fused data points is shown in the Figure 4 and we achieve a accuracy of more than 90% plotting the trajectory of target.

The joined probability approach, discussed before, is compared with 3 multi-sensor data fusion algorithms: 1) Simple barycentre (SB) of raw measurements [15] 2) Weighted detection quality barycentre of raw measurements (WDQB) [16] 3) PDA algorithm: Weighted barycentre with the observation to track probability [17] and further Kalman filter (PDAKF) application [18]. Four different metrics, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Computational Time, and Detection accuracy, are considered to compare these three methods with the proposed method. As shown in Figure 5a, the PDAKF algorithm has the lowest RMSE of 1.4 m, closely followed by the proposed approach with an RMSE of 1.5 m. MAE metric is very similar to the RMSE metric with PDAKF and the proposed approach having the lowest error, and it is shown in Figure 5b. In both the RMSE and MAE metrics, the SB and WDQB approaches showed the highest error, which makes the proposed approach a good alternative.

Regarding computation time, SB and WDQB approaches outperform the proposed algorithm, shown in Figure 5c. The PDAKF algorithm is the most computationally intensive, with a time of 1 second. Detection accuracy is shown in Figure 5d, and again, the PDAKF algorithm outperforms all other methods with 96% accuracy. However, the proposed approach has a detection accuracy of 95%, which is very competitive. The other two approaches, SB and WDQB, show an accuracy of 85% and 88%, respectively.

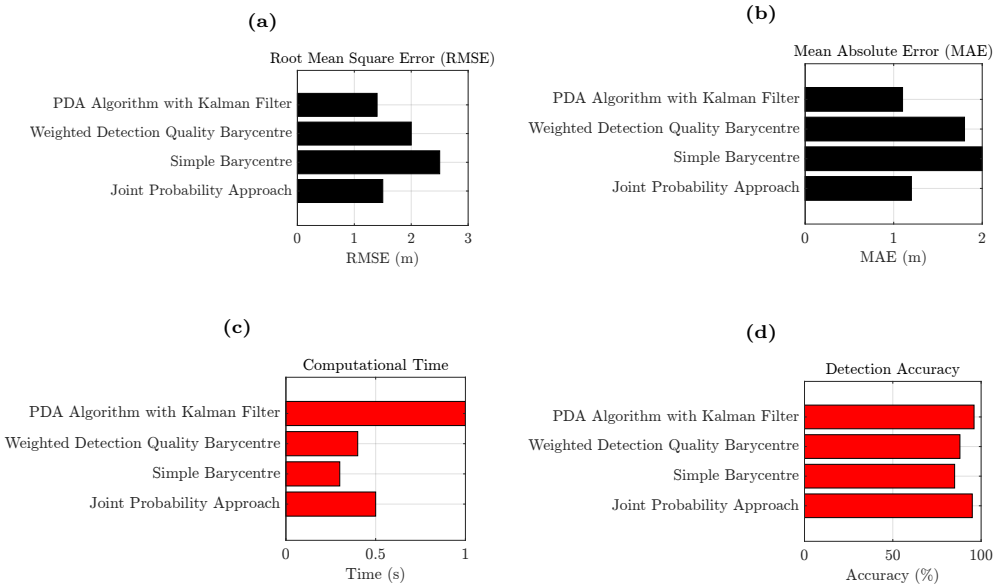


Fig. 5. Comparison of data fusion methods

Overall, analysis indicates that the Joint Probability Approach offers a balanced performance with high accuracy and relatively low computational complexity. Although the PDAKF algorithm shines in each metric, on the other hand it is significantly more computationally demanding.

5 Conclusion

The Joint Probability Approach (JPA), with its competitive accuracy and efficiency, provides a robust alternative for multi-sensor data fusion, making it suitable for real-time applications where both accuracy and computational efficiency are critical. This article shows that the numerical error obtained with the JPA is in the same order as the PDA algorithm. However, the first approach's computational load and complexity are lower. The proposed approach can be applied to fuse data from these sensors, providing a comprehensive and accurate representation of the drone's position and trajectory. By combining the strengths of various sensors, JPA can enhance the detection reliability and reduce false alarms, which are crucial for effective drone monitoring and control. Future work could further focus on optimizing the JPA computational aspects to enhance its applicability in various real-time scenarios. Moreover, this method can be combined with emerging sensor technologies and Artificial Intelligence-based analytics for improved performance. Overall, the Joint Probability Approach can be made even more effective for drone detection and C-UAV applications, providing a reliable and efficient solution for modern security challenges.

Acknowledgements. The principal investigator of the data fusion algorithm is Prof. Walter Matta (email: w.matta@unilink.it).

The work of this manuscript is also partially supported by NATO Emerging Security Challenges Division Science for Peace and Security Programme Ai4CUAV – Innovative AI-framework to enable the Detection, Classification and Tracking of Killer-Drones SPS.MYP.G6246 signed 22 November 2024.

References

[1] Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A.: Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. IEEE Geoscience and Remote Sensing Magazine **7**(1), 6–39 (2019) https://doi.org/10.1109/MGRS.2018.2890023

[2] Barbedo, J.G.A.: Data fusion in agriculture: Resolving ambiguities and closing data gaps. Sensors **22**(6), 2285 (2022)

[3] Cataldo, D., Gentile, L., Ghio, S., Giusti, E., Tomei, S., Martorella, M.: Multibistatic radar for space surveillance and tracking. IEEE Aerospace and Electronic Systems Magazine **35**(8), 14–30 (2020)

[4] Ahuja, B., Gentile, L., Marotrella, M.: Improving Satellite Position and Velocity Calculation During Low Thrust Maneuvers Using Multi-Bistatic Radar and Unscented Kalman Filter. In: 4th IAA Space Situational Awareness Conference. IAA, Florida (2024)

[5] Dong, J., Zhuang, D., Huang, Y., Fu, J.: Advances in multi-sensor data fusion: Algorithms and applications. Sensors **9**(10), 7771–7784 (2009)

[6] Hong, T., Liang, H., Yang, Q., Fang, L., Kadoch, M., Cheriet, M.: A real-time tracking algorithm for multi-target uav based on deep learning. Remote Sensing **15**(1), 2 (2022)

[7] Yang, F., Xu, F., Fioranelli, F., Le Kernec, J., Chang, S., Long, T.: Practical investigation of a mimo radar system capabilities for small drones detection. IET Radar, Sonar & Navigation **15**(7), 760–774 (2021)

[8] Lupidi, A., Forti, A.C., Jajaga, E., Matta, W.: An artificial intelligence application for a network of lpi-fmcw mini-radar to recognize killer-drones. In: WEBIST, pp. 320–326 (2022)

[9] Jajaga, E., Rushiti, V., Ramadani, B., Pavleski, D., Cantelli-Forti, A., Stojkovska, B., Petrovska, O.: An image-based classification module for data fusion antidrone system. In: International Conference on Image Analysis and Processing, pp. 422–433 (2022). Springer

[10] Kumar, A., Giusti, E., Mancuso, F., Ghio, S., Lupidi, A., Martorella, M.: Three-dimensional polarimetric inisar imaging of non-cooperative targets. IEEE Transactions on Computational Imaging **9**, 210–223 (2023)

[11] Kumar, A., Giusti, E., Mancuso, F., Martorella, M.: Polarimetric interferometric isar based 3-d imaging of non-cooperative target. In: IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 385–388 (2022). IEEE

[12] Kumar, A., Giusti, E., Martorella, M.: Hybrid polarimetry inverse sar. In: 2023 IEEE International Radar Conference (RADAR), pp. 1–6 (2023). https://doi.org/10.1109/RADAR54928.2023.10371162

[13] Azimirad, E., Haddadnia, J., Izadipour, A.: A comprehensive review of the multi-sensor data fusion architectures. Journal of Theoretical & Applied Information Technology **71**(1) (2015)

[14] Thomopoulos, S.C.: Sensor integration and data fusion. Journal of Robotic Systems **7**(3), 337–372 (1990)

[15] Achaji, L., Daher, M., El Najjar, M.E.B., Charpillet, F.: Multi-sensor data fusion for smart home reliable pedestrian localization. In: 2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET), pp. 144–149 (2021). IEEE

[16] Sara, D., Mandava, A.K., Kumar, A., Duela, S., Jude, A.: Hyperspectral and multispectral image fusion techniques for high resolution applications: A review. Earth Science Informatics **14**(4), 1685–1705 (2021)

[17] Macagnano, D., Destino, G., Abreu, G.: A comprehensive tutorial on localization: Algorithms and performance analysis tools. International Journal of Wireless Information Networks **19**, 290–314 (2012)

[18] Chen, B., Zhang, W., Hu, G., Yu, L.: Networked fusion kalman filtering with multiple uncertainties. IEEE transactions on Aerospace and Electronic Systems **51**(3), 2232–2249 (2015)

[19] Bernardo, J.M., Smith, A.F.: Bayesian Theory vol. 405. John Wiley & Sons, ??? (2009)

[20] Ribeiro, R.A., Falcao, A., Mora, A., Fonseca, J.M.: Fif: A fuzzy information fusion algorithm based on multi-criteria decision making. Knowledge-Based Systems **58**, 23–32 (2014)

Utilizing Vision Large Language Models for Automatic Image Annotations: A Comparative Study

Ali Abd Almisreb[1[0000-0001-7581-5747]], Tarik Namas[1[0000-0003-0254-9382]] , Özge Büyükdağlı[1[0000-0001-5758-4607]] , Alessandro Cantelli-Forti [2[0000-0002-6943-2632]] , Edmond Jajaga[3[0000-0003-1833-5856]] and Nurlaila Ismail[4[0000-0002-8504-4875]]

[1]Faculty Computer Science and Engineering, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina
[2]Lab RaSS CNIT, Pisa, 56124 Pisa
[3]Mother Teresa University, Mirçe Acev nr. 4, 1000 Skopje, North Macedonia
[4]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
aalmisreb@ius.edu.ba

Abstract. Image annotations can be a time-consuming task. This study looks into how well the OWLv2 and Grounding-DINO-Tiny models can annotate objects in four categories: airplanes, birds, drones, and helicopters. We revealed the preliminary results or findings as follows by comparing the confidence scores and the detection rate. The Grounding-DINO-Tiny model was quite successful, offering no empty frames and relatively high confidence scores most of the time for the distinguished categories such as the helicopter and drone. Still, it fared poorly in birds, having lower confidence scores or more annotations with a value less than 50% which signifies the model's weakness in identifying birds. The proposed model, OWLv2, had fairly moderate outcomes and the quality of data differed from one category to the other which undermined the reliability of the model. For the enhancement of the future performance, there are several recommendations that we make; these include; improving the ability to identify birds, eliminating inconsistency in the datasets, and improving on the quality of the data gathered.

Keywords: image annotation, OWLv2, Grounding-DINO-Tiny.

1 Introduction

The occurrence of large numbers of image and video data as central aspects in computer vision and machine learning necessitates the need for fast and accurate data annotation. Originally, most of the annotation have always been done manually and there is a time effect and inconsistency that has led to the need to look for an automatic way. New-generation deep learning and natural language processing enable the creation of the Vision Large Language Models (VLLMs) that can be considered as a viable solution for the automation of data annotation (Chen et al., 2022; L. Zhou et al., 2020). VLLMs are trained on large quantities of image-text pairs, thus, allowing them to learn the correlation between the visual characteristics of images and their linguistic descriptions.

Such capability enables VLLMs to not only produce precise descriptions of the image but also, important, semantically relevant (Vaswani et al., 2017).
Because of the described characteristics, VLLMs have a great prospect to become the key to accurate and scalable data annotation. The need to advance this area of study is evident when considering the growing use of large-scale annotated collections in numerous AI-related fields. Thus, through analyzing the advantages and limitations of the existing VLLMs, this study seeks to help advance the methods of annotation, thus promoting the advancement of the field of computer vision.

This study aims to:

- Evaluate OWLv2 and Grounding-DINO-Tiny VLLMs' performance on data Annotating tasks.
- Compare the accuracy, efficiency, and consistency of VLLM-generated annotations.
- Provide analysis of certain advantages and disadvantages of VLLM usage depending on the kind of the data and the type of the annotations.
- Provide a guideline that would describe the usage of effective VLLM during the data annotation process.

When accomplishing the identified objectives, this article aims at asserting the VLLMs' value as an instrumental tool for annotation. Besides, it seeks to offer significant findings that will be useful for future research in this specialty and for future developments.

The remainder of this article is structured as follows: Section 2 provides background of the topic. Section 3 provides a detailed overview of VLLMs and their application in data annotation. Section 4 presents the methodology employed in this study. The results and evaluation are presented in Section 5, followed by the conclusion and future work in Section 6.

2 Background

Large Language Models explain the current generation of Artificial Intelligence systems designed for analyzing and synthesizing texts similar to human languages. These models are trained with code and text data and involves the usage of deep learning, especially transformers. This training assists them in comprehending the nation's syntactical, spoken and written English – both its patterns and meanings. Therefore, LLMs are highly skilled in a range of natural language processes like the ones that involve text generation, language translation, content summarization, and question answering (Brown et al., 2020; Devlin et al., 2019). Some of these are OpenAI's GPT-4 (OpenAI, 2024a) and Google's Gemini (Gemini Team et al., 2023) are among the notable masters, which show the competency of how these models generate meaningful and contextually relevant text patterns.

The Vision Large Language Models (VLLMS) are the superior versions of the LLMs which is primarily designed to close the apparent gap between the visual and text data (Islam et al., 2024; Jiang et al., 2024; Y. Zhou et al., 2024). These VLLM models are quite capable of incorporating the features of visual and language data and thus enabling them to facilitate the comprehension and description of the content of the visuals in natural language. Hybrid models, like VLLMs, which combine Computer Vision and Natural Language Processing (NLP are capable of performing tasks such as generating descriptions, for images answering questions related to content and even analyzing visuals and telling stories (Lu et al., 2019; Radford et al., 2021a). Top notch examples consist of OpenAIs CLIP, by (Radford et al., 2021b). DALL-E by (OpenAI, 2024b) showcasing progress, in comprehending multimodal information.
Data labeling is a task, in machine learning when creating supervised models. It involves tagging or annotating data to establish the ground truth. Examples of computer vision tasks falling into this realm include object detection, image segmentation and landmark identification, within images (Emam et al., 2021; Miceli et al., 2020; Mullen et al., 2019). Creating reliable annotations is crucial, for developing resilient models as they significantly impact the models capacity to understand and apply insights, from the data. Datasets such, as ImageNet (Deng et al., 2009) and COCO (Lin et al., 2014) have played a role in pushing forward research in computer vision. However the manual annotation process is labor intensive, costly and susceptible to mistakes. This highlights the importance of automated annotation techniques with VLLMs showing potential as a solution due, to their ability to offer precise annotations (Tan et al., 2024).

3 Vision LLMs for Data Annotation

To perform the process of data annotation, Vision Large Language Models (VLLMs) engage in the optical and the linguistic comprehension. The VLLMs models can work with the images and output corresponding labels since Image Recognition is paired with Natural Language Processing. The detection what is the object is implemented with the help of image recognition, and language processing ability for providing the appropriate label. The over all procedure usually takes three procedural steps. The first process that happens is the feature extraction by the CNNs or vision transformers to get features which are objects, textures, or spatial relations of the image (Dosovitskiy et al., 2021). The second process is the recognition of the context and the meaning of the picture (Radford et al., 2021). The third step is to introduce comments on the annotations which have to be made by using the knowledge of features, context and semiotics of the input image. The annoations can be utilized for indicating the objects, explaining what is inside the picture or providing an adequate title for the picture (Lu et al., 2019).
The VLLMs have done many annotation tasks with a high degree of automation where the human labor of manual annotation has been substantially minimized or eliminated by such tasks as object detection (Redmon & Farhadi, 2017), segmentation (Ronneberger et al., 2015), image captioning (Anderson et al., 2018), and visual question answering (Antol et al., 2015; Zhang et al., 2024).

4 Methodology

In this study, we utilized a dataset collected by (Svanström et al., 2020), the dataset contains videos of four types of flying objects: airplanes, birds, drones, and helicopters You can classify by types of vehicles and many more depending with the need. These videos were converted into the separate frames. Next, the frames were passed through two VLLM models. The first model is OWLv2 which is the VLLM derived in the work of (Minderer et al., 2023), however, it is more advanced as it combines vision and language to enable tasks like object detection. And the second model is Grounding Dino Tiny model (Liu et al., 2023) Although, the overall architecture of this Grounding Dino Tiny model is distinct, it is an open-set object detector that combines the power of DINO Transformer-based detectors with state-of-art grounded pre-training. This integration allows the model to pick out objects that are denoted by the human language inputs, which is remarkable in the approach. Both of the models were trained to make outputs in the YOLO format. This study applied a range of methods, which is illustrated in the following figure; Figure 1: Methodology.

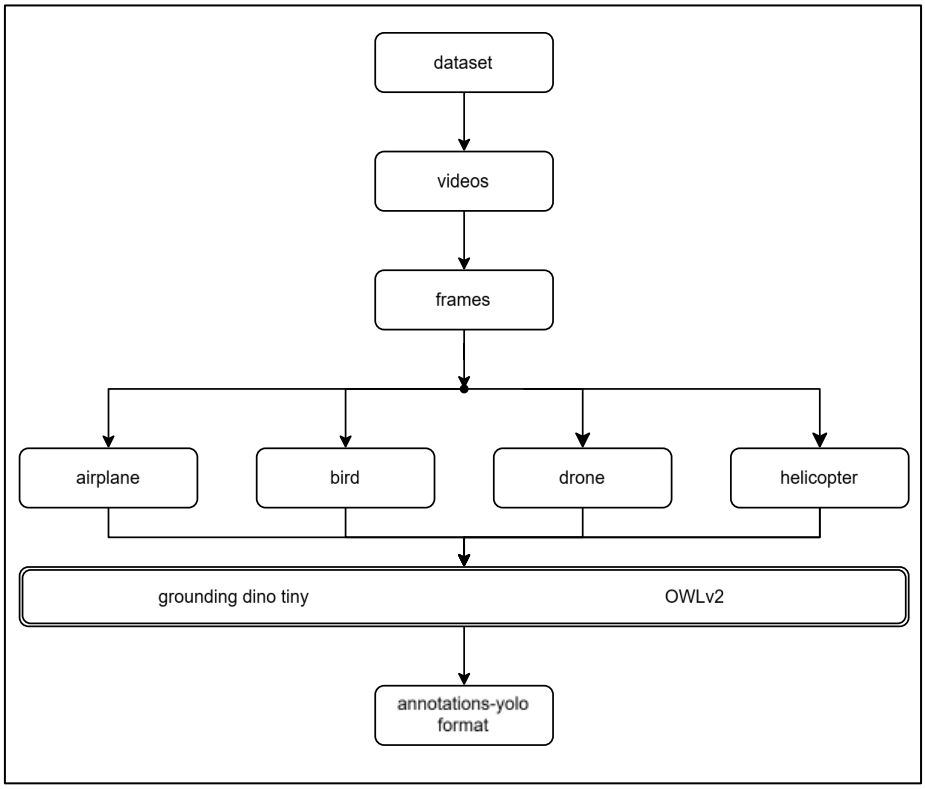


Fig. 1: methodology steps

Mathematically, this process can be represented as follows:

- Let I be the set of input frames.
- For each frame ($i \in I$):
  - $D = \text{model}(\text{processor}(i, \text{class\_names}))$
  - where D includes bounding boxes B, scores S, and labels L.
- For each detected object $d \in D$:
  - Compute the YOLO format coordinates:

$$x_{\text{center}} = \frac{x_1 + x_2}{2 \cdot \text{image\_width}}$$

$$y_{\text{center}} = \frac{y_1 + y_2}{2 \cdot \text{image\_height}}$$

$$\text{width} = \frac{x_2 - x_1}{\text{image\_width}}$$

$$\text{height} = \frac{y_2 - y_1}{\text{image\_height}}$$

- Save the annotations $\left(x_{\text{center}ij}, y_{\text{center}ij}, \text{width}_{ij}, \text{height}_{ij}, s_{ij}\right)$ in the output directory,
  where, s_{ij} is the confidence score for the detected object j in image i.
- Image Annotation
  - draw bounding boxes and labels on the image i based on the converted coordinates:

$$\text{Annotated Image}_{\text{i}} = \text{draw}(\text{i}, \text{B}_{\text{i}}, \text{L}_{\text{i}}, \text{S}_{\text{i}})]$$

- Evaluation Metrics:
  - After processing all images, calculate the following metrics:
  - Total Images Processed:
    - $\text{T} = |\text{I}|$, where $|\text{I}|$ is the total number of images in the dataset.
  - Empty Frames:
    - $\text{E} = \sum_{\text{i} \in \text{I}} \delta(\text{D}_{\text{i}} = \emptyset)$ where δ is the Kronecker delta function, returning 1 if the detection result D_{i} is empty, and 0 otherwise.
  - Frames with Detected Objects:
    - $\text{F} = \text{T} - \text{E}$
  - Average Confidence Score:

- $\bar{S} = \frac{\sum_{i \in I} \sum_{j \in D_i} s_{ij}}{\sum_{i \in I} |D_i|}$, where s_{ij} is the confidence score for detected object j in image i, and $|D_i|$ is the number of detected objects in image i.

5 Results and Evaluation

This section presents sample of the annotated frames by OWLv2 and Grounding DINO tiny models. This is followed by critcal analysis of the performance of both models. The expermental eniveronment includes Python 3.12.4 perfomed on GPU NVIDIA Geforce RTX 2060 Super, equibed with 8 GB of memory. The system RAM is 32 GB.

5.1 Results

Figures 2 and 3 illustrate representative samples of annotated images generated by the OWLv2 and Grounding DINO tiny models, respectively. The visualizations highlight the performance and accuracy of each model in object detection and annotation tasks.

Fig. 2: annotated frames by OWLv2

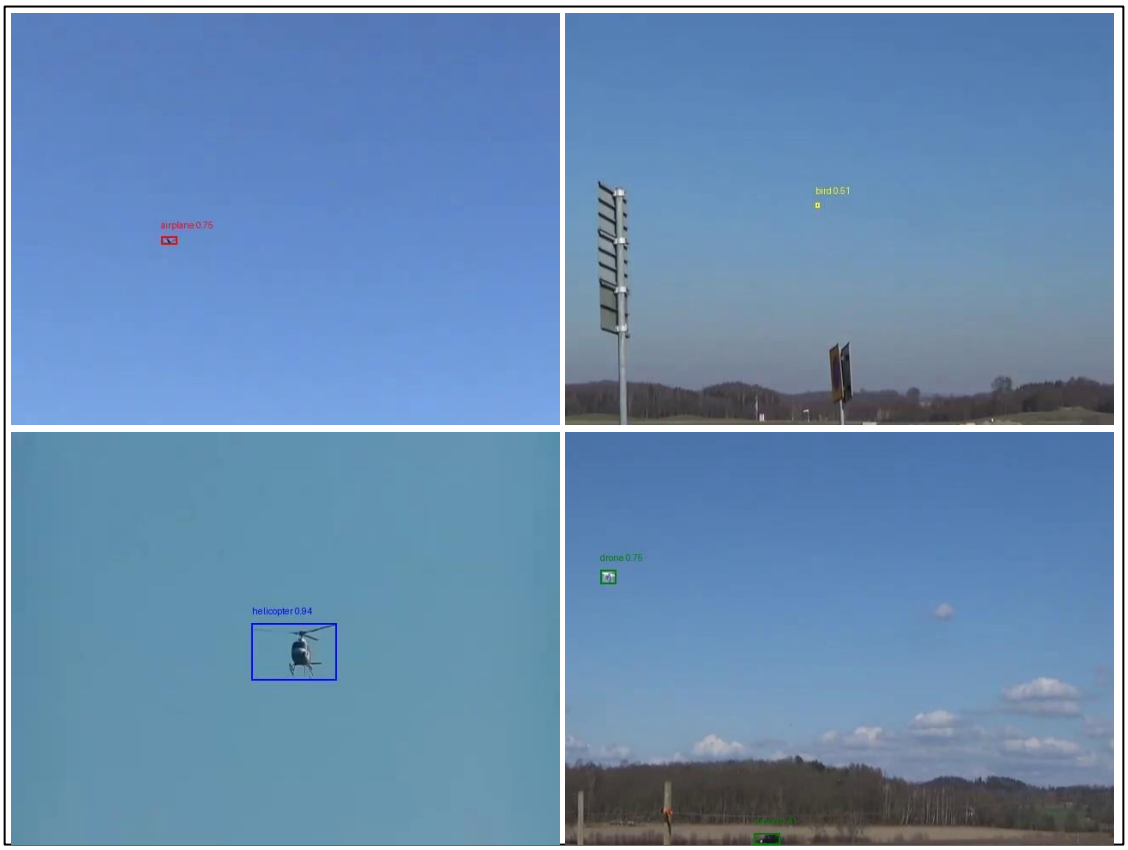


Fig. 3: annotated frames by Grounding DINO tiny

5.2 OWLv2 Performance Evaluation

The performance evaluation of OWLv2, as in Table 1, across different classes indicates notable variability in detection accuracy: Helicopter detection demonstrates the highest confidence with an average score of 0.670, showcasing the model's strong capability, likely due to an abundance of data for this class. Airplane detection exhibits moderate accuracy, with an average score of 0.619, indicating reasonable performance but room for improvement compared to helicopters. Bird detection records a lower average score of 0.409, highlighting challenges due to the complexity and variability in the appearance of birds. Drone detection is the most challenging, with the lowest average score of 0.329, suggesting significant difficulty in identifying drones, possibly due to their smaller size and fewer training examples. The occurrence of empty frames, where no objects are detected, also varies across classes: Helicopter and airplane classes show

relatively fewer empty frames, indicating a high detection rate and effective identification in most frames. In contrast, the drone class has a substantial number of empty frames, with 408 out of 3998 images containing no detected objects, underscoring detection difficulties. Similarly, the bird class has a considerable number of empty frames, with 225 out of 5494 images, confirming the detection challenges reflected in the confidence scores.

Table 1: OWLv2 performance evaluation

Class	Total Images	Empty Frames	Frames with Objects	Average Confidence Score
Airplane	8449	179	8270	0.619
Bird	5494	225	5269	0.409
Drone	3998	408	3590	0.329
Helicopter	13759	155	13604	0.670

The critical analysis of the model's performance based on the object detection confidence score as in Table 2 reveals varied results across different categories. For airplanes, the majority of annotations (36.06%) fall within the 91-100% confidence range, indicating strong detection capabilities, although 11.55% are in the 21-30% range, showing occasional struggles. In contrast, bird detection shows significant difficulty, with 36.89% of annotations within the 0-20% confidence range, highlighting the need for improvement. Drone detection is particularly challenging, with over half of the annotations (52.42%) in the 21-30% range, indicating inadequate model performance and dataset quality. Helicopter detection, however, is robust, with 33.92% of annotations in the 71-80% range and 27.82% in the 81-90% range, demonstrating strong detection capabilities.

Table 2: Object Detection Confidence Analysis for OWLv2 model

Class	Total Files	Total Anno-tations	0-20%	21-30%	31-40%	41-50%	51-60%	61-70%	71-80%	81-90%	91-100%
Airplane	8449	8596	0 (0.00%)	993 (11.55%)	734 (8.54%)	690 (8.03%)	691 (8.04%)	884 (10.28%)	1503 (17.49%)	3099 (36.06%)	2 (0.02%)
Bird	5494	15069	0 (0.00%)	5560 (36.89%)	2735 (18.15%)	2130 (14.13%)	1963 (13.02%)	1657 (11.00%)	996 (6.61%)	28 (0.19%)	0 (0.00%)
Drone	3998	3590	0 (0.00%)	1882 (52.42%)	875 (24.38%)	465 (12.95%)	241 (6.71%)	112 (3.12%)	15 (0.42%)	0 (0.00%)	0 (0.00%)
Helicopter	13759	14564	0 (0.00%)	990 (6.80%)	842 (5.78%)	850 (5.84%)	986 (6.77%)	1896 (13.02%)	4941 (33.92%)	4051 (27.82%)	8 (0.06%)

5.3 Grounding-DINO-Tiny Performance Evaluation

The critical analysis of the Grounding-DINO-Tiny model, based on the contents of Table 3, shows no empty frames, indicating its effectiveness in detecting objects across all images. In terms of confidence scores, the helicopter class has the highest average score (0.785), followed by drone (0.727), airplane (0.752), and bird (0.476), suggesting the model is most confident in detecting helicopters and least confident with birds. The model demonstrates consistency in detection, as no empty frames were observed across the dataset. High confidence scores for helicopters and drones indicate reliable detection capabilities for these classes, while the lower confidence score for the bird class (0.476) suggests possible overfitting or issues with the training data for this class.

Table 3: grounding DINO tiny model performance evaluation

Class	Total Images Processed	Empty Frames	Frames with Detected Objects	Average Confidence Score
Airplane	8449	0	8449	0.752
Bird	5494	0	5494	0.476
Drone	3998	0	3998	0.727
Helicopter	13759	0	13759	0.785

In Table 4, the critical analysis of the Grounding-DINO-Tiny model performance based on the object detection confidence score, for the Airplane category, the majority of annotations (59.14%) fall within the 81-100% confidence range, indicating a high level of annotation certainty, with only 19.55% of annotations below 50% confidence, suggesting a generally reliable dataset with a need for slight improvements. For the Bird category, a large proportion of annotations (36.10%) are within the 21-30% confidence range, and 50.06% are below 50% confidence, indicating a significant portion of annotations have lower confidence levels, impacting the reliability of the data and requiring focused improvements. The Drone category shows that the majority of annotations (73.90%) are in the 81-90% confidence range, suggesting high reliability, with only 12.51% of annotations below 50% confidence, indicating a highly trustworthy dataset with strong detection capabilities. For the Helicopter category, most annotations (67.84%) are above 70% confidence, with the highest concentration (54.50%) in the 91-100% range, and only 22.04% of annotations below 50% confidence, suggesting a high degree of confidence in the annotations and robust detection performance.

Class	Total Files	Total Annota-tions	0-20%	21-30%	31-40%	41-50%	51-60%	61-70%	71-80%	81-90%	91-100%
Airplane	8449	9855	0 (0.00%)	803 (8.15%)	366 (3.71%)	459 (4.66%)	693 (7.03%)	821 (8.33%)	884 (8.97%)	1765 (17.91%)	4064 (41.23%)
Bird	5494	19338	0 (0.00%)	6981 (36.10%)	2648 (13.69%)	2372 (12.27%)	2758 (14.26%)	2058 (10.64%)	1543 (7.98%)	919 (4.75%)	59 (0.30%)
Drone	3998	4507	0 (0.00%)	305 (6.77%)	128 (2.84%)	54 (1.20%)	54 (1.20%)	68 (1.51%)	257 (5.70%)	3331 (73.90%)	310 (6.88%)
Helicopter	13759	15869	0 (0.00%)	703 (4.43%)	691 (4.35%)	722 (4.55%)	888 (5.60%)	1268 (7.99%)	836 (5.27%)	2117 (13.34%)	8644 (54.50%)

6 Challenges and Limitations

Even here, the Grounding-DINO-Tiny model produces notably lower confidence scores on the bird samples compared to those of the other classes; a portion of them is below 50% in both the training and valid-ation datasets; this could mean that birds are hard to identify due to the variety of their appearances or sampling issues in the DINO. While the param-eter values yields no empty frame, and high average confidence scores for most of the classes, one might observe a likely sign of over fitting particularly in the bird class for the training and validation datasets. In the case of OWLv2 model, several classes are associated with lower quality annotations except for class 'bird' as it sheds half of the annotations quality which in turn affects the reliability of the data set. Verifying the quality of the data and data selection, the creation of a diverse and large dataset is a protection of the proper improvement of the model. Even though the Grounding-DINO-Tiny model performs well in detecting helicopters and drones, it has comparatively less confidence when it comes to airplanes and birds, which tells that there are drawbacks regarding the detection because of the variation in the appearances and aspects linked to these classes. As these two models perfrom well with visible images, they would not show promising results based on infra-red images.

7 Future Work

To optimize bird class detection, one has to enhance the number of the bird images in the training data set as well as apply methods of data augmentation to obtain more diverse training samples. Other ways to improve the model for bird detection are also possible, for instance, tuning hyperparameters or employing transfer learning with a dataset that is more specialized for bird detection. Improving global dataset quality is needed to make training and validation full-featured, and non-reducing in terms of generalized models overfitting, and therefore implementing cross-validation approaches. Improving the overall quality of data entails stringent quality checks on the annotations to achieve high confidence scores for each class and employing sophisticated annotations tools and strategies in order to achieve accurate annotations.

8 Conclusion

Evaluating the OWLv2 and Grounding-DINO-Tiny models, we observe that both show reliable detection and annotation results for helicopters as well as drones. On the other hand, for bird class it is a more challenging task because of the lower confidence scores and large number of annotations having less than 50% confidence pointing to area where targeted improvement can be made. Improving the quality of data, including providing more balanced and representative datasets or annotations higher in state-quality VLLMs with advanced architectures can greatly enhance performance consistency across all detection applications. Solving these issues could then lead to more accurate

and better image annotation in every category, increasing the usefulness of this stunning models on implementation.

References

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00636

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual Question Answering. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2425–2433. https://doi.org/10.1109/ICCV.2015.279

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *2020-December*.

Chen, J., Guo, H., Yi, K., Li, B., & Elhoseiny, M. (2022). VisualGPT: Data-efficient Adaptation of Pretrained Language Models for Image Captioning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2022-June*. https://doi.org/10.1109/CVPR52688.2022.01750

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *CVPR09*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE. *ICLR 2021 - 9th International Conference on Learning Representations*.

Emam, Z., Kondrich, A., Harrison, S., Lau, F., Wang, Y., Kim, A., & Branson, E. (2021). *On The State of Data In Computer Vision: Human Annotations Remain Indispensable for Developing Deep Learning Models*. 139. https://arxiv.org/abs/2108.00114v1

Gemini Team, Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., Lillicrap, T., Lazaridou, A., … Vinyals, O. (2023). *Gemini: A Family of Highly Capable Multimodal Models*. http://arxiv.org/abs/2312.11805

Islam, M. S., Rahman, R., Masry, A., Laskar, M. T. R., Nayeem, M. T., & Hoque, E. (2024). *Are Large Vision Language Models up to the Challenge of Chart Comprehension and Reasoning? An Extensive Investigation into the Capabilities and Limitations of LVLMs*. http://arxiv.org/abs/2406.00257

Jiang, Y., Yan, X., Ji, G.-P., Fu, K., Sun, M., Xiong, H., Fan, D.-P., & Shahbaz Khan, F. (2024). *Effectiveness assessment of recent large vision-language models*. https://doi.org/10.1007/s44267-024-00050-1

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*(PART 5). https://doi.org/10.1007/978-3-319-10602-1_48

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2023). *Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection*. http://arxiv.org/abs/2303.05499

Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in Neural Information Processing Systems*, *32*.

Miceli, M., Schuessler, M., & Yang, T. (2020). Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2). https://doi.org/10.1145/3415186

Minderer, M., Gritsenko, A., & Houlsby, N. (2023). *Scaling Open-Vocabulary Object Detection*. http://arxiv.org/abs/2306.09683

Mullen, J. F., Tanner, F. R., & Sallee, P. A. (2019). Comparing the Effects of Annotation Type on Machine Learning Detection Performance. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 855–861. https://doi.org/10.1109/CVPRW.2019.00114

OpenAI. (2024a). *ChatGPT [Large language model]*.

OpenAI. (2024b). *DALL-E: Introducing outpainting, https://openai.com/blog/dall-e-introducing-outpainting*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021a). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of Machine Learning Research*, *139*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021b). *Learning Transferable Visual Models From Natural Language Supervision*. http://arxiv.org/abs/2103.00020

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021c). *Learning Transferable Visual Models From Natural Language Supervision*. http://arxiv.org/abs/2103.00020

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*. https://doi.org/10.1109/CVPR.2017.690

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*. https://doi.org/10.1007/978-3-319-24574-4_28

Svanström, F., Englund, C., & Alonso-Fernandez, F. (2020). Real-time drone detection and tracking with visible, thermal and acoustic sensors. *Proceedings - International Conference on Pattern Recognition*. https://doi.org/10.1109/ICPR48806.2021.9413241

Tan, Z., Beigi, A., Wang, S., Guo, R., Bhattacharjee, A., Jiang, B., Karami, M., Li, J., Cheng, L., & Liu, H. (2024). *Large Language Models for Data Annotation: A Survey*. https://arxiv.org/abs/2402.13446v2

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009.

Zhang, J., Huang, J., Jin, S., & Lu, S. (2024). Vision-Language Models for Vision Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2024.3369699

Zhou, L., Palangi, H., Zhang, L., Hu, H., Corso, J. J., & Gao, J. (2020). Unified vision-language pre-training for image captioning and VQA. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i07.7005

Zhou, Y., Cui, C., Rafailov, R., Finn, C., & Yao, H. (2024). *Aligning Modalities in Vision Large Language Models via Preference Fine-tuning*. http://arxiv.org/abs/2402.11411