



MINISTRY
OF EDUCATION AND SCIENCE
OF UKRAINE



ACADEMY

OF TECHNICAL SCIENCES OF UKRAINE

CONNECTIVE
technologies

DEPARTMENT OF

*INFORMATION
TECHNOLOGY*

2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE)

Conference Proceedings

November 29-30, 2023
Ivano-Frankivsk, Ukraine

*Issued by the decision of the Academic Board of the Faculty of Mathematics and Computer Science
(protocol #10, 2023-11-21)*

This book and its papers are published under the Creative Commons License Attribution 4.0 International
(CC BY 4.0)

Editors

Mykola Kuz  and **Mykola Kozlenko** 

Department of Information Technology
Faculty of Mathematics and Computer Science
Vasyl Stefanyk Precarpathian National University
57 Shevchenka Street
Administration Building, Office #319
Ivano-Frankivsk, 76018, Ukraine
Tel: +380 (342) 59-60-58
E-mail: icisse@pnu.edu.ua

Papers are presented in authors' edition. The authors are responsible for the content of their paper. The organizers, committees, reviewers, or editors reserve the right not to be responsible for the topicality, correctness, completeness or quality of the information provided.

M. Kuz and M. Kozlenko, Eds., 2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE), Ivano-Frankivsk, Ukraine: Vasyl Stefanyk Precarpathian National University, 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.10397356>. Accessed: Nov. 29, 2023.

ISBN 978-966-640-549-7

DOI 10.5281/zenodo.10397356

The high-quality refereed papers appearing in this book compose the proceedings of the 2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE). ICISSE is a forum organized by the Department of Information Technology of Vasyl Stefanyk Precarpathian National University. 2023 2nd ICISSE is held from November 29 to 30, 2023 in Ivano-Frankivsk, Ukraine. The event is intended to bring together researchers, scientists, and engineers to discuss experimental and theoretical results, mainly in the area of software engineering. The conference also covers topics related to computer science, computer engineering, systems analysis, cybersecurity, information systems and technology, industrial automation, electronics, metrology, micro and nanosystems, telecommunications, radio frequency engineering, IT entrepreneurship, and IT education. The conference is included into the list of scientific conferences scheduled for 2023 by the Ministry of Education and Science of Ukraine.

Recommended paper citation template:

J. Author1, K. Author2, and L. Author3, "Decapitalized title of the paper", *2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE)*, Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, Nov. 29-30, 2023, pp. xxx-xxx.

ISBN 978-966-640-549-7

© Copyright for the individual papers by the papers' authors, 2023

© Copyright for the book as a collection by its editors, 2023

© Vasyl Stefanyk Precarpathian National University, 2023

COMMITTEES

General Chair:

Prof. Dr. Mykola Kuz

Technical Program Committee Chair:

Assoc. Prof. PhD Mykola Kozlenko

Local Organizing Committee Chair:

Mariia Dutchak

Publicication Chair:

Assoc. Prof. PhD Mykola Kozlenko

Preface

2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE) is a forum organized by the Department of Information Technology at Vasyl Stefanyk Precarpathian National University. The university has more than 80 years of history. Its inspirational success story was written by people who devoted themselves to science, education, and culture.

This year, ICISSE was held from November 29 to 30, 2023 in Ivano-Frankivsk, Ukraine. The event is intended to bring together researchers, scientists, and engineers to discuss experimental and theoretical results, mainly in the area of software engineering. The conference also covers topics related to computer science, computer engineering, systems analysis, cybersecurity, information systems and technology, industrial automation, electronics, metrology, micro and nanosystems, telecommunications, radio frequency engineering, IT entrepreneurship, and IT education. The conference is included into the list of scientific conferences scheduled for 2023 by the Ministry of Education and Science of Ukraine.

General Statistics: Submissions 77, Accepted 49, Acceptance rate 0.63.

Statistics by Country

country	authors	submitted	accepted	acceptance rate
Algeria	3	2.00	2.00	1.00
India	21	7.00	2.00	0.29
Morocco	3	1.00	1.00	1.00
Saudi Arabia	2	1.00	1.00	1.00
Sweden	1	0.25	0.25	1.00
Ukraine	115	63.75	41.75	0.65
United States	1	1.00	1.00	1.00

The organizers are grateful to participants for wonderful presentations and presence at the event.

TABLE OF CONTENTS

Paper ID	Paper	Page
1	Multi-Robot Control System Architecture Concept for Industrial and IoT Applications. Volodymyr Nazarenko	1
2	Multimodal Radar Data Fusion for Human Activity Recognition. Djazila Souhila Korti, Zohra Slimane	4
3	Exploring the Potential of UWB Radar in Human Activity Recognition: A Brief Survey. Djazila Souhila Korti, Zohra Slimane	11
7	Cloud giants: AWS, Azure and GCP. Igor Nevludov, Svitlana Sotnik	18
8	Адаптивний захист інформації на мобільному пристрої. Євген Бровченко, Валерій Самарай, Володимир Павленко	24
11	Entropy, Gaussian Distribution and Fractional Processes. Anatoliy Malyarenko, Yuliya Mishura, Kostiantyn Ralchenko and Yevheniia Anastasiia Rudyk	31
12	Застосування автоматизованих засобів для забезпечення прийняття рішень при віддаленому управлінні. Олег Посашков, Олександр Цимбал	35
13	Intelligent decision support system based on recurrent neural networks and genetic algorithm for stock trading. Dmytro Uhryn, Andrii Bilyk	38
14	Використання гнучких комутаційних структур у складі апаратної частини мобільного робота. Ігор Невлюдов, Ірина Жарікова, Артем Бронніков	41
15	Метод аналізу супутникових знімків на основі глибинного навчання для застосування в обороні країни. Юрій Ушенко, Дмитро Угрин, Владислав Дашкевич	45
16	Analysis of a fixed-width binning method. Olga Solovei	49
17	Структура матриць системи лінійних алгебраїчних рівнянь для задачі моделювання масопереносу в пористому середовищі на графі. Валерій Колесников	53
18	Розробка автоматизованої системи розрахунку та прогнозування показників екологічного стану. Аліна Ольвінська, Вікторія Рувінська	57
19	Розробка системи розділення джерел звуку на основі методів машинного навчання. Орест Ткачук	60
21	Simulation of critical phenomena in the island systems in general relativity with flows. Hennady Shapovalov, Anatoly Kazakov, Vyacheslav Oleynyk	67
22	Розробка та оптимізація нейронної мережі для автоматичного визначення об'єктів на зображеннях за допомогою глибинного навчання. Данило Рудяга	73
23	Дослідження переваг використання фреймворку Nuxt.js у створенні сучасних веб-додатків. Дмитро Угрин, Артем Карачевцев	77
24	Математична модель інтелектуального обробника запитів. Ольга Тузенко, Наталія Сідун	82
25	A Novel Approach for Power-Efficient Voltage Level Shifting in Digital Circuits. Vakulabharanam Ramakrishna, Alenoor Krishna Kumar, Sistla.V.Sudheer Kumar	88
26	Comparison of versions of the YOLO algorithm for recognizing violations of individual labor protection rules in the workplace. Olha Pronina, Olena Piatykor	94
29	Sentiment Analysis for Student Feedback. Halyna Melnyk, Vasyl Melnyk	99
31	Обертальний криптоаналіз деяких функцій ускладнення ARX-криптосистем. Сергій Яковлев, Денис Кобець	101
32	Підходи до автентифікованого шифрування. Наталія Щур	105
35	Real-time big data analysis systems resulting from the Internet of Things IoT. Mohammed A. Makarem, Muneef A. Razaz	110
36	Anomaly Detection Techniques in Communication and Network Systems. Lesia Mochurad, Ivan Dubravskyi	126
37	Формування цільової функції визначення оптимальної зарядки літій-іонних акумуляторів. Сергій Буряк, Оксана Гололобова	134
38	Modeling of Wave Propagation in Dispersive Media Using New ADE-TLM Method. El Hadi El Ouardy, Hamid Bezzout, and Hanan El Faylali	138
40	Створення тренувальних датасетів для корекційних великих мовних моделей. Володимир Нестеренко	147

41	Використання машинного навчання для управління процесами аграрної економіки. Петро Грицюк, Максим Гаврилюк	150
42	Identify the perpetrator in the gathering Using AI. Soumya Upadhyay, Vishal Chauhan, Vaibhav Sharma, Yashveer Makhloga, Madhav	155
43	Застосування Grasshopper Optimization Algorithm при навчанні нейронних мереж. Андрій Ляшкевич, Анастасія Дейнеко, Юлія Шевчук	162
44	Теорія та практика реалізації базових функцій інтелектуальної АСУ в нейромережевому базисі. Сергій Альошин, Олена Гайтан	167
45	Використання інформаційної системи для формування індивідуальної освітньої траєкторії. Сергій Шаров	174
46	Набуття soft skills IT фахівцем в процесі онлайн навчання. Ольга Чуб, Марина Новожилова	179
47	Порівняння ефективності застосування технологій штрихового кодування та RFID у логістичних процесах. Ігор Невлюдов, Андрій Слюсар, Софія Хрустальова, Кирило Хрустальов, Віктор Косенко	183
49	Мобільний застосунок для вимірювання земельних ділянок. Ігор Мерлак	191
51	Інтелектуальне управління міською інформаційною інфраструктурою. Наталія Братерська	195
54	Методи автоматизації та оптимізації побудови навчального матеріалу в інтелектуальних адаптивних вебсистемах самоосвіти. Марія Дутчак, Андрій Аннич, Олег Козич	198
56	Сертифікаційна модель автоматизованої програмної системи управління навчальним закладом WEBportal ПНУ. Руслан Запухляк, Микола Кузь, Микола Козленко, Микола Пікуляк, Ігор Лазарович, Валерій Ткачук, Борис Незамай	203
57	Дослідження асинхронних методів сервісної взаємодії у веб-додатках. Микола Пікуляк, Станіслав Домбровський	207
58	Системний аналіз аквапонних систем. Роман Залозний, Наталія Заєць	214
60	Розробка методології автоматизованого тестування мобільних застосунків. Владислав Остражнов	218
62	Modeling a method for generating a stream of secret keys in the form of permutation matrices for encryption-masking of video frames and studying its characteristics. Vladimir Krasilenko, Vladislav Podlubnyi, Diana Nikitovich	222
67	Methods and algorithms for forecasting the effective use of energy resources for household consumers. Roman Dyndyn, Ihor Lazarovych, Serhii Ishcheriakov	232
68	Вибір нефункціональних вимог при розробці програмного забезпечення. Володимир Кімак	237
70	Моделювання ліній розділу фільтраційних потоків за умови відсутності перетоку із джерелом збурення. Сергій Каштан	242
71	Підвищення надійності системи логування. Іван Савка, Тарас Іванишин	247
75	Монотонні предикати в теорії ігор. Оксана Микицей	251
77	Weak sinusoidal signal extraction from white noise using convolutional neural network. Mykola Kozlenko	254

Multi-Robot Control System Architecture Concept for Industrial and IoT Applications

Volodymyr Nazarenko

*Department of Computer systems, networks and cybersecurity
National University of Life and Environmental Sciences of Ukraine
Kyiv, Ukraine*

Abstract—This article is about proposing new robot configuration by introducing a new planning algorithm. The scope of the paper research is the process of sequential part assembly in a multi-robot environment. The focus is robot collaboration and proposition of optimal robot configuration mechanism to overcome common problems.

Keywords—*Computer Science, IoT, Robotics, Computer Vision, SLAM, Swarm, System Architecture.*

I. INTRODUCTION

The scope of this research proposal is limited to general software and system model concept. The paper focuses on finding robot base and arm configuration for optimal task performance while overcoming existing constraints. The main existing status quo constraints is that robots cannot intersect with one another (having sequence of operations, it has to predict future states) [1]. When moving to the next sequential operation robot has two possibilities: 1) change grasp position on the part; 2) move part to next operation while maintaining same grip position. While humans can perform both tasks simultaneously, robots must be programmed to do one of them. This presents a problem for teams of robots to decide and which task to perform. We will be examining this issue and how it can be solved withing this paper.

II. RELATED WORK

The existing solutions mostly focus on and revolve around multi-robot grasp planning as Constraint Satisfaction Problem (CSP); with each grasp position in each operation being a variable [2]. This variable imposes two types of constraints: collision and transfer. In order to be able to locate optimal solution to robot grasp planning paper presents solution algorithm, while assuming that robot should preform regrasp between each of operations. This enables us to solve problems as a set of smaller CSP, which can be solved separately for each assemble operation [3]. The advantage of the commonly used algorithm is that anytime planner can be stopped at each given time stamp of the computational process. With it the time of computation combined with the cost of removing excessive regrasp operation gets larger than cost of performing certain singular specific operation. As input existing algorithm takes sequence of relative positions of assembly parts. The problem which CSP helps to solve is – find grasping configurations for all robots required by the assemblies in all operations [4].

III. METHODOLOGY

The goal of this research is to study and examine typical industrial or IoT application use case of multi-robot configuration system. The aim is to look for optimal problem solution that allows for a minimal number of robots regrasp operations. For this purpose, we consider a typical manufacturing process. This process is carried out as a set of sequential assembly operations, each operation being

either change grasp or move part to next operation without changing the grip [5]. To solve CSP problem we examine two solutions: 1) backtracking search (worst case scenario – exponential in number of variables); 2) local neighborhood computation (solve conflicts in local neighborhood of the graph). The proposed algorithm starts from solving the easiest problem first: solve constraint graph with no transfer constrains, then gradually add new ones and solves them. The flow of the algorithm is 1) solve problems for each of connected components using Back Tracing search; 2) set collection of solutions from sub operations (step1) as current best solution; 3) start incremental addition of new transfer constraints, reducing number of regrasps (this procedure tries to solve problems as fast as possible, iterating over all valid combinations, prioritizing this solutions, in order to try to solve them first, this reduces time as we conduct search only in local neighborhood of graph).

IV. RESULTS

Results of observed experimental data show that proposed algorithm concept works efficiently and better than direct – brute-force computation. It generates the first “best” solution in approx. 4 seconds, in most cases optimal result required one regrasp operations. We believe that the optimal case is to study robots’ cooperation and communication systems with reference to exiting communication software architecture. As an example, we can use client-server architecture, where robots that assemble product are clients, and we have Computational cluster with server running on it. The communication between server and client is wireless, this requires examination possible interferences to the network through various factory production elements. Going further in this theme we can implement Swarm intelligence (derived from UAV) to robots and have robots exchange information in real time between each other. This approach requires development of design making process how and when robot can decide about his next set of action, while cooperating with others. This will help to solve the problem when robots find themselves trapped inside the structure it is assembling.

Also, in real world factory settings there are usually additional constraints: limited space for robot navigation conveyor belts or moving parts [4]. Robots can be unable to place parts for regrasp operations or they may encounter new obstacles while moving part to next production chain. This setting requires additional improvements to the CSP algorithm.

As mentioned in the paper, in future we must consider robots that are used to insert fasteners. They can be added as new variable to the equation, which can have several states: wait, fasten, and move. For algorithm testing and evaluation such robots can either be stationary or they can be mobile (on platforms). An example of stationary robot with arm that have only one degree of freedom vertical or horizontal, to make it easier for other robots to bring parts for fastening operation. Also, this robot can have a tracking laser to improve tracking. All this presents new challenges to the CSP algorithm. We can add new variables which show robot current state or used to determine future states. This is useful if robot can carry different tasks, It can be added to SolveTransferConstraints and used during computation of new best solution, serving as transition between one point on the graph and next one (next operation).

Besides we can add the number of robots to the equation, to calculate optimal number of robots for the task. If robot can have to state either hold or insert fasteners, we can determine, which one of the groups should be doing first task, which ones second. Of course, we must limit the number of robots, have redefined maximum, minimum number for specific task, but equation will calculate optimal number.

V. DISCUSSION

Robots have limited computational power, also if we take energy efficiency into account proposed architecture can be helpful for future iterations of the multi-robot control system architecture for

industrial and IoT applications. For example, if we go further in the topic of robots' motion planning, then robots should have better localization or perception systems.

VI. FUTURE RESEARCH

In future we plan to implement SLAM, to efficiency compute we need to gather all the information from robots' sensors pass it to computational unit which makes decision for them and will assign unique task to each robot, improve their cooperation algorithm on the run. This can be useful, if for instance, robots are responsible for several tasks during the assembly process. IN case of single robot proposed approach allows it to carry out several unique tasks, even just fastening bolts and bringing parts to the assembly. Secondly, this can improve algorithm computational time even further, as we could use it in real time with low latency. Robots can have some small computational units which have limited abilities but can make small decisions on their own. While exchanging sensors and other information to server which can send this robot to different task or send additional l robots to help.

VII. CONCLUSION

The area of industrial and manufacturing robotics control systems is a complex multidisciplinary field of study. To present a useful algorithm and system architecture of multi-robot control system architecture for industrial and IoT applications we need to create a test set-up, that includes real world physical robot manipulators, sample field typical use-case set-up. The presented algorithm and system model are useful steppingstones for future work on finding optimal multirobot control model.

REFERENCES

- [1] Chen, J. Y., & Barnes, M. J. (2014). Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44(1), 13-29.
- [2] Brailsford, S. C., Potts, C. N., & Smith, B. M. (1999). Constraint satisfaction problems: Algorithms and applications. *European journal of operational research*, 119(3), 557-581.
- [3] Ryan, M. (2010, May). Constraint-based multi-robot path planning. In *2010 IEEE International Conference on Robotics and Automation* (pp. 922-928). IEEE.
- [4] Song, P., & Kumar, V. (2002, May). A potential field based approach to multi-robot manipulation. In *Proceedings 2002 IEEE International Conference on Robotics and Automation* (Cat. No. 02CH37292) (Vol. 2, pp. 1217-1222). IEEE.
- [5] Dogar, M., Spielberg, A., Baker, S., & Rus, D. (2019). Multi-robot grasp planning for sequential assembly operations. *Autonomous Robots*, 43, 649-664.

Multimodal Radar Data Fusion for Human Activity Recognition

Djazila Souhila Korti, Zohra Slimane

*Department of Telecommunication, Faculty of Science and Technology, SSL Laboratory
Belhadj Bouchaib University
Ain-Temouchent, Algeria*

Abstract— The growing interest in the use of multi-sensor data fusion technology is driven by the need for accurate and reliable human activity recognition (HAR), which surpasses the capabilities of single-sensor technology. In this study, a deep learning model is introduced to combine data from multiple sensor systems and automatically extract features for human activity classification. This is achieved using a public database comprising three datasets containing micro-Doppler signatures of human activities. These datasets include frequency-modulated continuous waves (FMCW) at 77 GHz, frequency-modulated continuous waves at 24 GHz and ultra-wideband radio pulses (IR-UWB) at 10 GHz. The model's performance is evaluated using different fusion strategies, including early, halfway and decision fusion. The results indicate that halfway fusion is the most efficient approach for multi-sensor fusion-based HAR, achieving the highest accuracy of 95.41%.

Keywords—*Human activity recognition, multi-sensor, radar, multimodal, fusion, deep learning.*

I. INTRODUCTION

Human activity recognition (HAR) is of paramount importance across diverse domains, including healthcare, safety, and assistive technologies [1], [2]. Traditional HAR methods heavily rely on single-source data, such as data from sensors like radars, accelerometers and cameras [3], [4]. However, the advent of cutting-edge technologies has heralded a transformative era of multimodal learning, which significantly enhance the precision and effectiveness of HAR systems [5]-[7].

Multimodal learning is rooted in the concept of fusing rich data from multiple sensors [8]. By integrating data from a variety of sources, researchers seek to obtain a holistic view of human activities. This process extracts complex elements that clearly characterize human movement. Thanks to this approach, HAR systems are able to provide more accurate, comprehensive and contextual recognition of human activities. This is particularly useful in areas where accuracy is paramount, underlining the importance of multimodal learning in the field of HAR.

On the other hand, radar sensors offer a unique perspective in HAR [9]-[11]. This approach goes beyond conventional wearable sensors and cameras, providing a non-intrusive and versatile method for monitoring human activities. Radar sensors emit radio waves and detect their reflections, allowing them to capture detailed movement patterns without physical contact, ensuring privacy preservation and user-friendliness. Furthermore, radar-based HAR systems are not affected by environment conditions or obstructions, making them suitable for a wide range of environmental scenarios.

Research in HAR using radar technology have developed numerous techniques to make activity identification more accurate [12]. These techniques involve various deep learning methods to extract valuable insights from extensive datasets. Some common methods used in this context include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Hybrid models. Hybrid models are gaining prominence because they combine different learning approaches. Typically, these models use CNNs to capture spatial information and merge them with

other models like Long Short-Term Memory networks (LSTMs), Support Vector Machines (SVM), and Extra Trees (ETs) [13]-[15]. This fusion of techniques makes hybrid models a powerful tool for improving the precision of HAR.

The main objective of this research study is to investigate different data fusion methodologies for radar-based multimodal HAR systems. To achieve this objective, we have chosen to adopt three distinct approaches, namely: 1) early fusion, 2) halfway-fusion and 3) decision fusion. The study explores an alternative method for training a multiple-input hybrid model using a publicly available database. Through a comprehensive analysis, we aim to discover the most effective way of fusing data from multiple radar sources for accurate recognition of human activities.

II. METHODOLOGY

A. Dataset

We used the multi-frequency RF sensor network's human activity database provided by Gurbuz *et al.* [5]. The database consists of three datasets acquired from three synchronized radar sensors operating in monostatic mode. The sensors include the Texas Instruments 77 GHz Frequency Modulated Continuous Wave (FMCW) radar tuned to a central frequency of 77 GHz with a bandwidth of 750 MHz, the Ancortek 24 GHz FMCW radar with a central frequency of 24 GHz and a bandwidth of 1500 MHz, and the Ultra-Wide Band (UWB) XeThru X4 impulse radar transmitting in a band of approximately 7 GHz - 10 GHz. During data acquisition, the sensors were placed side by side at a height of 1 meter above the ground. Six participants of different ages, sizes, and weights took part in this study. In total, 11 different activities were considered, namely WLKT (walking towards the radar), WALKA (walking away from the radar), PICK (picking up an object from the ground), BEND (bending over), SIT (sitting on a chair), KNEEL (kneeling), CRWL (crawling towards the radar), LIMP (limping with a stiff right leg), WTOES (walking on both toes), SHTEPS (walking with small steps), and SCSSR (walking with scissors). Each participant moved from 0.5 to 3 meters along the radar's line of sight and performed 10 repetitions of each activity. This resulted in a total of 60 samples per class and per radar. Fig. 1 represents the micro-Doppler spectrograms of different activities for each radar.

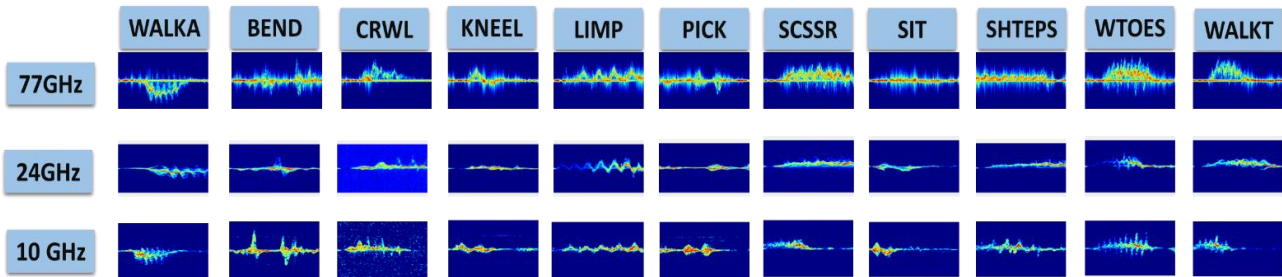


Fig. 1 Micro-Doppler signature for each radar/activity

B. Model description

In this work, we employ our previously published model known as the Multi-Input Multi-Output Convolutional Extra Trees (MIMO-CxT) [14]. This model has been specifically designed to accommodate multi-sensor systems and efficiently handle data from multiple sources. MIMO-CxT is a sophisticated fusion of CNN and the ETs algorithm. CNNs are adept at extracting spatial features from data, making them ideal for tasks like image analysis. On the other hand, the Extra Trees algorithm, a type of ensemble learning method, excels at making predictions based on complex data relationships. By merging these two approaches, MIMO-CxT can effectively process and analyze multi-source data from various sensors. The detailed structure of the MIMO-CxT is depicted on Fig. 2.

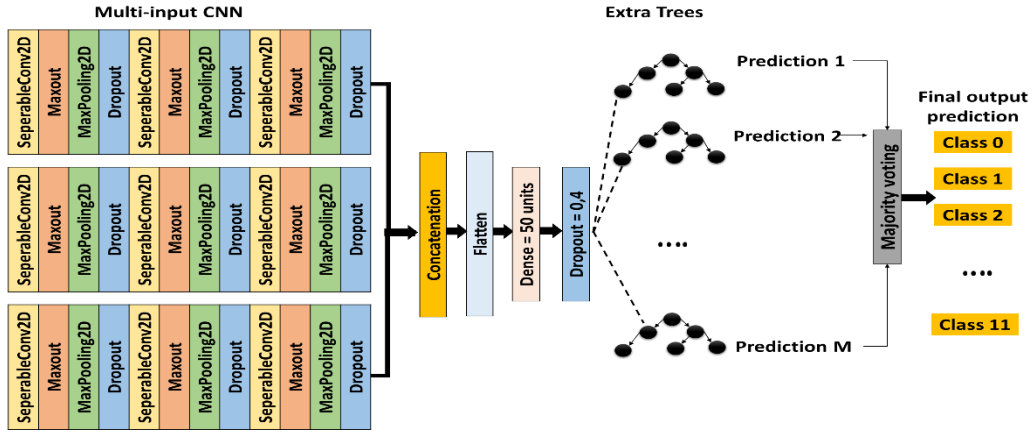


Fig. 2 MIMO-CxT architecture [14]

C. Fusion strategies

To assess the model's performance trained on multimodal datasets, we implemented three distinct fusion techniques as depicted in Fig. 3.

- Early Fusion: In the early fusion approach, we combine the sensor data from different frequencies at the beginning of the processing pipeline. This approach facilitates the incorporation of data from diverse sensors into a unified representation right from the beginning.

- Decision Fusion: In the decision fusion approach, data integration takes place after the individual processing of each sensor's output. This approach enables a more in-depth analysis of the data from each sensor. The combined information is then used to make informed decisions regarding human activity recognition.

- Halfway Fusion: The halfway fusion strategy strikes a balance between early and decision fusion. It involves merging data from multiple sensors at an intermediate stage of the processing, offering flexibility in feature extraction and integration.

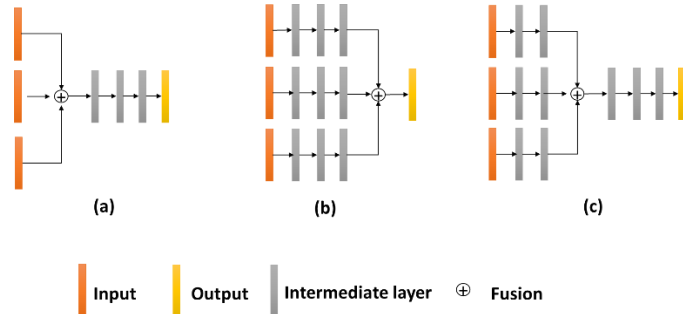


Fig. 3 Fusion strategy: (a) early, (b) halfway, (c) decision

D. Performance measure

In order to evaluate the performance of our proposed model, different metrics have been used including: confusion matrix, accuracy, precision, recall, and F1-score. These metrics are based on True Positives (T_P), True Negatives (T_N), False Positives (F_P), and False Negatives (F_N).

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \quad (1)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (2)$$

$$Recall = \frac{T_P}{T_P + F_N} \quad (3)$$

$$F1\text{-score} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

E. Implementation details

The model is implemented in Python using Keras framework with Tensorflow backend. Other libraries used include sklearn, Pandas, Numpy and Yellow brick. The hardware configuration is based on a running environment of Intel (R) Core (TM) i5@2.40 GHz CPU, 16GBs of RAM, 1To of hard disk and Windows 10. All the samples within the three datasets underwent a transformation, being resized to dimensions of 75×75 and then converted into binary images. This conversion to binary images serves several purposes, including faster inference, avoiding unnecessary pre-processing steps, and reducing storage requirements. Following this, the converted samples were subject to a random split, with 80% allocated for training and the remaining 20% for testing. To ensure consistency across experiments, a random seed parameter was employed for the selection of test samples. With the exception of replacing depth separable convolutional layers with conventional convolutional layers and changing the number of units in the last dense layer to 150, all the preset model configurations remain unchanged, as mentioned in [14].

III. RESULTS

The training procedure consists of two primary phases. Initially, the three-input CNN-Softmax undergoes training for 100 epochs using the back-propagation algorithm, with a batch size of 16. During this phase, it serves as a feature extractor and can follow one of three fusion strategies: early, halfway, or decision fusion. In the second phase, Softmax is substituted with the ETs classifier, which receives the features extracted and combined from the CNN branches.

The outcomes of individual experiments conducted under the chosen fusion strategy are presented below:

A. Early fusion

The confusion matrix and the classification report obtained from the early fusion strategy are depicted in Fig. 4 (A) and Fig. 5 (A), respectively.

B. Halfway fusion

The confusion matrix and the classification report obtained from the halfway fusion strategy are depicted in Fig. 4 (B) and Fig. 5 (B), respectively.

C. Decision fusion

The confusion matrix and the classification report obtained from the decision fusion strategy are depicted in Fig. 4 (C) and Fig. 5 (C), respectively.

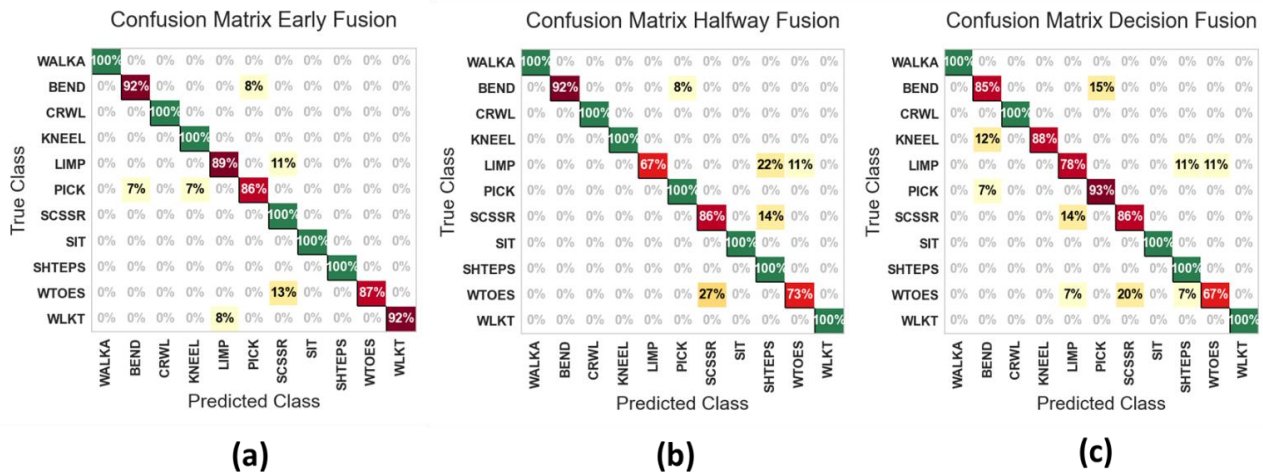


Fig. 4 Confusion matrix: (a) early, (b) halfway, (c) decision fusion

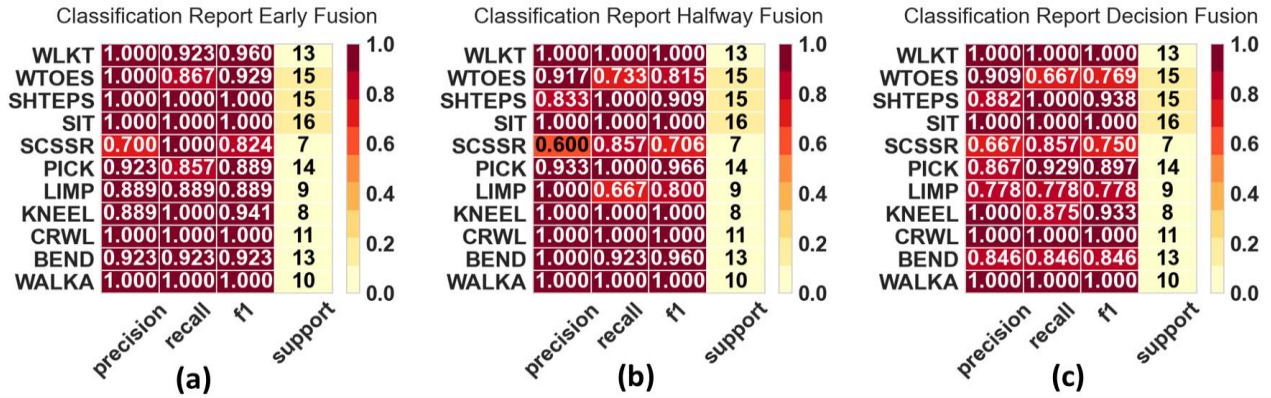


Fig. 5 Classification report: (a) early, (b) halfway, (c) decision, fusion

D. Comparison

- Fusion strategy

To evaluate the model's performance across the different fusion strategies, we have assessed several key parameters. These parameters encompass train accuracy, test accuracy, runtime prediction, and the number of parameters of the CNN. The comparative analysis of performance is presented in Table 1.

Table 1. Comparative analysis of performance for fusion strategy

Fusion strategy	Train acc (%)	Test acc (%)	Runtime prediction (s)	CNN N° of parameters
Early fusion	100	94.65	0.230664	87059
Halfway fusion	100	95.41	0.264935	254099
Decision fusion	100	90.83	0.443204	257699

- Previous work

The results of our model can be directly compared to the results of Mahbubur *et al.* [6], as this work used the same dataset used in the experiments of this study. Table 2 summarizes the performances of the models.

Table 2. Performance comparison

Model	Accuracy
Cross model DNN [6]	93%
MIMO-CxT	95.41%

IV. DISCUSSION

When we evaluate the performance of a model using early, halfway and decision fusion strategies, it's clear that the choice of fusion method has a significant impact on model efficiency. Table 1 highlights that the halfway fusion strategy excels, achieving the highest accuracy at 95.41%. In this approach, individual modalities are initially processed independently, allowing for the extraction of optimized features for each modality. The CNN branches are interconnected midway in the process, utilizing data from all three modalities to train the remainder of the network. This strategy not only capitalizes on correlations between data types but also reveals crucial patterns within each unimodal dataset that may go unnoticed with early fusion techniques.

Comparing early and decision fusion, early fusion achieves the highest accuracy at approximately 94.65%. It accomplishes this with a reduced number of training parameters and faster runtime predictions compared to decision fusion. Early fusion integrates multimodal information right at the start of data processing, conserving computational resources and ensuring the model leverages all available information from the beginning. This aids in creating a comprehensive understanding of the input, capturing intricate relationships and patterns that decision fusion might overlook. Additionally, it eliminates the need to store and process redundant data separately, saving both memory and processing power.

In summary, both early and halfway fusion techniques yield impressive accuracies. However, halfway fusion emerges as a holistic and efficient fusion approach that bridges the gap between early and decision fusion, showcasing its potential to enhance the analysis and classification of multimodal data

Furthermore, our approach allows for a direct comparison with the findings of Mahbubur *et al.* [6], as both models are trained on identical datasets without any parameter tuning. The MIMO-CxT provided significantly better performance and outperformed it by a rate of 2.41%. Notably, the MIMO-CxT boasts a lightweight architecture compared to the cross model DNN. This highlights the significant potential and interest in the combination of CNN and ETs, which leads to achieving outstanding performance.

V. CONCLUSION

In this work, an investigation was conducted to explore various fusion strategies for HAR using multimodal radar data. The findings revealed that the halfway fusion approach proved to be highly effective in combining features and information, ensuring that the model fully utilizes the available data. Additionally, the study demonstrated that the integration of CNN and ETs offers a notable performance improvement compared to using CNN alone. This suggests that the fusion of these technologies enhances the overall performance of HAR models, making them more effective in their tasks.

REFERENCES

- [1] G. Diraco, G. Rescio, P. Siciliano, and A. Leone, "Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing," *Sensors*, vol. 23, no. 11, p. 5281, 2023.
- [2] D. Nagpal and S. Gupta, "Human Activity Recognition and Prediction: Overview and Research Gaps," *IEEE 8th Int. Conf. Conver. Technol. I2CT*, 2023, pp 1-5.
- [3] F. M. Noori, M. Z. Uddin, and J. Torresen, "Ultra-Wideband Radar-Based Activity Recognition Using Deep Learning," *IEEE Access*, vol. 9, pp. 138132–138143, 2021.
- [4] N. Bento, J. Rebelo, A. V. Carreiro, F. Ravache, and M. Barandas, "Exploring Regularization Methods for Domain Generalization in Accelerometer-Based Human Activity Recognition," *Sensors*, vol. 23, no. 14, pp. 1–12, 2023.
- [5] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, and F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-Doppler signature classification (Rising Researcher)," vol. 11408, no. 2020, p. 16, 2020.
- [6] M. Mahbubur Rahman and S. Z. Gurbuz, "Multi-Frequency RF Sensor Data Adaptation for Motion Recognition with Multi-Modal Deep Learning," *IEEE Natl. Radar Conf. – Proc.*, 2021.
- [7] R. J. Piechocki, X. Wang, and M. J. Bocus, "Multimodal sensor fusion in the latent representation space," *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, 2023.

- [8] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous Human Activity Recognition With Distributed Radar Sensor Networks and CNN-RNN Architectures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
- [9] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, and A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition," *IEEE J. Microwaves*, vol. 3, no. 3, pp. 1–13, 2023.
- [10] J. Maitre and S. Bouchard, K. Gaboury, "Data filtering and deep learning for enhanced human activity recognition from UWB radars," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–12, 2023.
- [11] Z. Wang, A. Ren, Q. Zhang, A. Zahid, and Q. H. Abbasi, "Recognition of Approximate Motions of Human Based on Micro-Doppler Features," *IEEE Sens. J.*, 2023.
- [12] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, 2019.
- [13] H. Zhou, Y. Zhao, Y. Liu, S. Lu, X. An, and Q. Liu, "Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System," *Sensors*, vol. 23, no. 10, 2023.
- [14] D.S. Korti, Z. Slimane, "Unobtrusive hand gesture recognition for post stroke patient rehabilitation using ultra-wide band radar and deep learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no 6, pp. 6872-6881, 2023. available: <https://ijece.iaescore.com/index.php/IJECE/article/view/32272>.
- [15] S. Huan, L. Wu, M. Zhang, Z. Wang, and C. Yang, "Radar Human Activity Recognition with an Attention-Based Deep Learning Network," *Sensors*, vol. 23, no. 6, 2023.

Exploring the Potential of UWB Radar in Human Activity Recognition: A Brief Survey

Djazila Souhila Korti, Zohra Slimane

*Department of Telecommunication, Faculty of Science and Technology, SSL Laboratory
Belhadj Bouchaib University
Ain-Temouchent, Algeria*

Abstract—The rapid advancements in Artificial Intelligence, particularly Deep Learning, have generated a rising demand for the development of automated tools for analyzing human behavior using various sensor technologies. Human Activity Recognition (HAR) through Impulse Radio Ultra-Wide Band (IR-UWB) has recently emerged as a compelling area of research. It offers numerous benefits, such as safeguarding individual privacy and low energy consumption. IR-UWB-based HAR holds promise in diverse applications, including eldercare, rehabilitation, and patient monitoring. The primary aim of this study is to provide insight into the progress and potential of IR-UWB based HAR. This investigation delves into its applications, challenges, and the role of deep learning in enhancing recognition accuracy. To accomplish this, we conduct a brief survey of existing techniques and applications in the realm of HAR using IR-UWB, focusing on data representation, feature extraction, and activity classification methods.

Keywords—*Human activity recognition, radar, ultra-wide band, machine learning, deep learning.*

I. INTRODUCTION

Human Activity Recognition (HAR) is known as the automatic process of understanding people's behaviors and their interactions with real world environment. It becomes an active and emerging research area as it incorporates wide-ranging and potential applications uses in a variety of domains, including health-care monitoring [1], sport [2], smart home [3], Human-Computer Interaction (HCI) [4], virtual reality [5], surveillance and security [6].

An activity refers to the movement of the whole body and/or its limbs [7]. Based on the complexity, the duration, and the involved body parts used for motion, we can distinguish different categories of activities [7]:

- **Gesture:** defined as the elementary movement of only one human's body limbs. It is generally characterized by a low complexity and temporal duration.
- **Action:** defined as a simple activity, which combines a series of gestures, organized temporally, and include the movement of more than one body part.
- **Interaction:** involves the presence of two subject; one must be a human, while the other can be a human or an object. It could be a human-human interaction or a human-object interaction.
- **Activity:** often called an event, is a more complex bodily movement, which describes the gestures or actions of one or multiple persons. An activity can include:
 - a set of actions of a single person.
 - an interaction or a set of interactions of a person with one or more objects.
 - an interaction or set of interactions between several people and/or objects.

- Group activity: known as the most complex type of activity. It can be seen as a set of activities performed by groups of several people and/or objects.

A typical HAR system works in three basic steps, which involves: (1) Data collection; (2) Data preprocessing; (3) Activities classification. However, for the data collection different sensing technologies have been used, in order to extract accurate and opportune information based on people's activities. They can be divided into two major categories on the basis of the sensor's deployment, which are: i) wearable device and ii) device-free.

A significant amount of work has been done on activity recognition using wearable sensor-based device such as, accelerometer, gyroscope, light sensor, temperature sensor, magnetometer and smartphones [8]–[10]. However, these devices must be worn or attached to the user's body, while it performs any activity, which usually make it uncomfortable for 24/7 monitoring and feeling constrained [10]. Moreover, in some cases wearing these devices is not feasible, for example the case of elderly users, which usually forget to wear it [10], and premature infants due to skin sensitivity [11]. To tackle the above-mentioned problems, researchers focused on developing approaches based on device-free in which users are not required to carry any device with them. The main idea is to deploy the sensors in the environment, and capture through them the data while the subjects perform different activities [12].

Vision based systems, are widely applied in the field of HAR due to the advantage of simplicity, direct perceiving and large amount of information it can capture. They use images and video frames to distinguish human behaviors and movements recorded from cameras [13]. Despite the high accuracy that such systems offer for identifying human activities, it is difficult to achieve good results in some scenarios due to some limitations related to vision systems [13]. For example:

- Installed cameras in long distances and/or high places provide low-quality videos, where the subject do not appear clear.
- Inappropriate light conditions make the cameras fail to work, especially at night.
- The demand of line-of-sight, which is not always obvious.
- The presence of severe occlusions in large and crowded places.
- The non-preservation of the individual's privacy, when used in some places such as homes, or bedrooms.

Non-contact sensors could be more favorable to use as they offer solutions to many problems associated with wearable sensors and vision-based systems [14]. Impulse Radio Ultra-Wide Band (IR-UWB) has recently emerged as one of the effective and promising non-contact sensors for human activity recognition due to a number of advantages over the existing technologies [1]. IR-UWB radars provide a simple and inexpensive system [1]; they are able to perform well in both highly lit and dark environments due to the fact that they are not affected by the weather conditions and the different light intensities [1]. IRUWB can completely avoid the problem of occlusion as it offers the advantage of penetration through obstacles and walls as well as the personal privacy since it deals only with signals [1]. Deep Learning techniques have achieved tremendous success in the HAR field, with their ability to learn complex and hierarchical features automatically from IR-UWB data and make predictions with high accuracy and less preprocessing compared to traditional Machine Learning algorithms. Different models have been proposed including convolutional neural networks (CNN) [15], Residual Networks [16], Long Short-Term Memory (LSTM) [17], Gated Recurrent Unit (GRU) [18].

II. RELATED WORK

Researchers made significant contributions in the HAR field using IR-UWB, for different applications, such as fall detection [19], apnea detection [20], elderly monitoring [21], etc. However, a typical HAR system mostly follows three main phases including (1) Data collection, (2) Data preprocessing, (3) Activities classification.

1) Data collection:

This first phase is very crucial and must be done in an appropriate way, as the whole HAR process depends on the gathered data and their quality. Two important points should be taken in consideration: the placement and the number of radars to use [17]. Incorrect selection of these two latter may adversely affect the recognition performance. Due to lack of resources, researchers tend in some cases to use public dataset, which allows their work to go far beyond the initial results found by other researchers and contribute to the pace of scientific discovery [22], [23].

2) Data preprocessing:

The collected raw data are reflected signals containing clutter as well as information related to the motion. A filtering is required to eliminate the unwanted signals using different methods such as loopback filter [1], DC removal [24], and averaging method [25]. After this the waveforms are organized in a matrix form, which represents a sequence of range bins (Fast time), each associated to a specific measurement time (Slow time). This matrix can be converted to different representation format where useful features can be extracted for classification. The clutter removed (Fast Time)-(SlowTime) matrix, known as time-series, is the simplest representation because of the minimum processing it requires. It was successfully used for fall detection with high accuracy [16], [26], [27]. However, such a representation is not widely used in activity recognition because of the poor information it provides. For this reason, the 2D representation takes place, as it carries sufficient human activity information from different domains namely: range-time (RT), Doppler-time (DT), range-Doppler (RD), and Cadence Velocity Diagram (CVD) [28].

Range-time map is a two-dimensional matrix, obtained by performing the Fast Fourier Transform (FFT) along the fast time axis. It results a 2D representation, which offers range information varying in time between the human subject and the radar, in addition to the duration of the activity [29]. Erol *et al.* [29], focused on using distance information in a fall detection algorithm to distinguish a real fall from a sitting position to reduce the false alarm rate. It was deduced that a real fall has twice the range of a sitting position. Shao *et al.* [30], used range-time maps to classify seven predefined actions. The model was able to achieve good results. However, it was demonstrated that the accuracy decreases when the subject is not on the same line of sight of the radar.

Doppler-time map also known as micro-Doppler spectrograms (MDS) are obtained by performing FFT with short and overlapped time windows on the range matrix [19]. It offers discriminative information about the motion and micro-motions, such as the velocity of the different body's parts as legs and hands.

Range-Doppler map provides both range and velocity information of a moving target at a specific time. It offers also the ability to measure the distance between the human subject and the radar. Such a representation is very useful for applications like tracking and locating targets [31].

Cadence Velocity Diagram (CVD) is obtained by applying FFT to the micro-Doppler signatures, along the time axis for all Doppler bins. It shows useful features like the shape, the size, and the frequency of the curves in the DT domain. Moreover, the velocities repeat created by the moving body limbs of the target can be extracted [32].

However, using a single radar/domain representation is usually insufficient for activity recognition, for example, when the target changes its distance and orientation in the same time with respect to the radar line of sight. The data fusion is then required to improve the overall classification performance by combining complementary information from different radars/domains [17]. Jokanovic *et al.* [33], showed the effectiveness of the human activity classification based on a multi-domain representation, where range-time, Doppler-time and range-Doppler information were combined for falls detection. M. Mostafa, and S. Chamaani [24], used the combination of Doppler-time and Range-time for distinguishing between fall and non-fall actions with 96.9% of accuracy. Bouchard *et al.* [34] proposed a system for movement recognition in a 40 meters apartment. A set of UWB radars was required to achieve an 80% accuracy. Li *et al.* [17], fused different extracted features from Doppler-time map and CVD to classify both individual and sequential human gait using a Frequency Modulated Continuous Wave (FMCW) radar and three IR-UWB. The 3D representation combines the three variables: time, range and Doppler in a cube format. It can be obtained by gathering Doppler-time maps that vary according to range [35], or stacking range-Doppler maps as a function of time [36]. Despite the fact that it offers plentiful information about the activity, it is not widely used due to the low number of algorithms that directly process 3D data, and the complicated process used to extract significant features [13].

(3) Activities Classification:

Many approaches have been proposed in the literature, varying from traditional Machine Learning, which required handcrafted features extraction to end-to-end automatic Deep Learning classification approaches [28].

- Machine Learning in HAR:

Park *et al.* [37] extracted Principal Component Analysis (PCA) from the received IR-UWB signal to classify hand gestures. A neural network containing one hidden layer, with nine units and an output layer with six units corresponding to the number of gestures were trained on these features. The model achieved an accuracy of 100%. However, it was tested on a reduced number of gestures which is not very conclusive. Khan *et al.* [38] developed a five-hand gesture recognition algorithm using one IR-UWB. Only three features, namely: variance of the Probability Density Function (PDF) of the magnitude histogram, Time Of Arrival (TOA) variation and the frequency of the reflected signal were used in addition to distance information as input parameter to K-means clustering algorithm. Seifert *et al.* [39] obtained an average accuracy of 88%, using sum-of-harmonics modelling as features extracted from the Doppler-time representation to classify the different gait styles, namely: normal, pathological and cane-assisted gaits. G. Wang and Z. Zhu [40] performed the recognition of human activities on the basis of temporal features, namely: energy, mean, square mean, margin, kurtosis, skewness and variance, for training a decision tree. The results show a poor accuracy when the target moves with tangential and radial motion. Ding *et al.* [35] treated human activity recognition as a multi-stage classification problem. Different features were extracted depending on the nature of the activity. Range-time maps were used in the first stage to perform a binary classification of activities when the target is static or moving with an accuracy of 99%. In the second stage, empirical physical characteristics and PCA were used to extract features from Weighted Range-Time-Frequency Transformation (WRTFT) representation, in order to classify 12 activities. They obtained an accuracy of 94,4% using K-Nearest Neighbors (KNN) algorithm for the in-situ motion and 95,3% using Bagged Trees for the non-in-situ motion. Kumar

et al. [41] classified four actions, with an accuracy of 85.25% using Doppler-time maps converted to a vector of features, and used as input to a support vector machine (SVM). Li *et al.* [17] classified individual and sequential gaits, using a Bi-LSTM (Bidirectional Long Short-Term Memory). 47 features from the Doppler-time and 7 features from the CVD were extracted to train the model. An accuracy of 98,2% and 93% was obtained for the individual, and the sequential gaits respectively. Chowdhury *et al.* [42] proposed a real-time human activity recognition. An accuracy of 86.4% was obtained using the KNN algorithm, trained on features extracted using the 2D2D-PCA (Bi-dimensional Principal Component Analysis) algorithm. Despite the good performance of Machine Learning algorithms in the field of HAR, they are sometimes neglected because of the time consumption, the selection of the appropriate features, the low capacity to manage large amounts of data, and the accuracy of classification, which largely depends on the effect of the extracted and selected features.

- Deep Learning in HAR:

Recently, Deep Learning approaches have gained more attention, due to their ability to extract robust and convenient features automatically without any prior domain knowledge [28]. The multiple layers, which contain a deep learning model, allow it to perform many transformations on the data, in order to learn low, middle, and high-level features. For that, difficult tasks such as HAR could be easily solved. Sadreazami *et al.* [19] used Doppler-time maps to train capsule network for fall detection. The hierarchical learned features make the model able to recognize with high accuracy of 95,66%, activities at different distances from the radar. Chen *et al.* [43] used the WRTFT representation as an input to a Deep Convolutional Neural Network (DCNN) for the recognition of six human activity. The results show a good generalization with an accuracy of 92.8%. Piriyaajakonkij *et al.* [25] developed Sleep PoseNet, a DCNN for different postures recognition during sleep. The model was trained on both time-difference and WRTFT representation an attained a 73.4% of accuracy. The authors suggested to use it for sleep disorders detection. Du *et al.* [44] proposed a Segmented Convolutional Gated Recurrent Neural Network (SCGRNN), which is a combination of CNN and GRU. The CNN layers were used to extract micro-Doppler spectrograms features. Based on these latter the GRU were used to calculate activity classification scores. The model was able to achieve 88.19% of accuracy. Yang *et al.* [18] showed the feasibility of using range profile for human motion recognition through the wall. They combined an Auto-Encoder Network (AEN) with three dense layers, with a GRU network with two hidden layers. The model achieved a 93% recognition accuracy within the initial 20% duration of the activities. The work presented by Ahmed and Cho [45] consisted of converting IR-UWB radar signal into grayscale images, then mapped this latter into (3D) RGB images for hand gesture recognition. An inception module was trained on these RGB images, and showed great results of about 95% accuracy. Khan *et al.* [46] implemented a system based on multiple IRUWB, able to recognize alphanumeric characters based on the pattern drawn in air using a CNN. Skaria *et al.* [47], presented the IR-UWB signature as a 3-dimensional tensor consisting of range-Doppler frames, and fed as input to CNN-based classifiers for hand gesture classification. Four different architectures were compared, and the CNN-LSTM achieved the highest accuracy of 96,15%. Hamidreza *et al.* [15] designed a fall detection method, able to reach a high accuracy of 98.37% based on time-frequency analysis. The Doppler-range maps were processed to obtain binary images and used as inputs to a CNN.

III. CONCLUSION

HAR serves as a fundamental step in understanding and interpreting human behavior. This review offers a concise exploration of established methodologies centered around activity recognition using IR-UWB. It delves into the fundamental elements shared by recognition systems, encompassing data representation, feature extraction, and activity classification. By shedding light on these key components, this review contributes to the comprehensive understanding of the HAR process, demonstrating its significance in a variety of fields.

REFERENCES

- [1] F. Khan, A. Ghaffar, N. Khan, and S. Ho Cho, "An Overview of Signal Processing Techniques for Remote Health Monitoring Using Impulse Radio UWB Transceiver," *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, 2021, doi: 10.1109/JPROC.2021.3100811.
- [2] A. Nadeem, A. Jalal, and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," 2021.
- [3] D. Bouchabou, S. M. Nguyen, C. Lohr, B. Leduc, and I. Kanellos, "A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning," 2021.
- [4] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, "Hand gestures recognition using radar sensors for human-computer-interaction: A review," *Remote Sens.*, vol. 13, no. 3, pp. 1–24, 2021, doi: 10.3390/rs13030527.
- [5] N. Zhang, T. Qi, and Y. Zhao, "Real-Time Learning and Recognition of Assembly Activities," pp. 1–15, 2021.
- [6] S. Mekruksavanich and A. Jitpattanakul, "Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models," *Electron.*, vol. 10, no. 3, pp. 1–21, 2021, doi: 10.3390/electronics10030308.
- [7] A. Benmansour, A. Bouchachia, and M. Feham, "Multioccupant activity recognition in pervasive smart home environments," *ACM Comput. Surv.*, vol. 48, no. 3, 2015, doi: 10.1145/2835372.
- [8] A. Ometov et al., "A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges," *Comput. Networks*, vol. 193, no. April, p. 108074, 2021, doi: 10.1016/j.comnet.2021.108074.
- [9] R. Abdel-Salam, R. Mostafa, and M. Hadhood, "Human Activity Recognition Using Wearable Sensors: Review, Challenges, Evaluation Benchmark," *Commun. Comput. Inf. Sci.*, vol. 1370, pp. 1–15, 2021, doi: 10.1007/978-981-16-0575-8_1.
- [10] R. Liu, A. A. Ramli, H. Zhang, E. Henricson, and X. Liu, "An Overview of Human Activity Recognition Using Wearable Sensors: Healthcare and Artificial Intelligence," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12993 LNCS, pp. 1–14, 2022, doi: 10.1007/978-3-030-96068-1_1.
- [11] S. Munidasa, P. Baghaei, E. Shim, O. Lin, and E. Ghafar-Zadeh, "A Bedsheet for Baby Monitoring at Night: Measurement and Characterization Results," *Can. Conf. Electr. Comput. Eng.*, vol. 2020-August, pp. 15–18, 2020, doi: 10.1109/CCECE47787.2020.9255671.
- [12] S. Bian, M. Liu, B. Zhou, and P. Lukowicz, "The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey," *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124596.
- [13] D. Nagpal and S. Gupta, "Human Activity Recognition and Prediction: Overview and Research Gaps," *2023 IEEE 8th Int. Conf. Conver. Technol. I2CT 2023*, no. September, 2023, doi: 10.1109/I2CT57861.2023.10126458.
- [14] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, 2019, doi: 10.3390/rs11091068.
- [15] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless Fall Detection using Time-Frequency Analysis and Convolutional Neural Networks," *IEEE Trans. Ind. Informatics*, vol. 3203, no. c, 2021, doi: 10.1109/TII.2021.3049342.
- [16] H. Sadreazami, M. Bolic, and S. Rajan, "Residual network-based supervised learning of remotely sensed fall incidents using ultra-wideband radar," *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2019-May, pp. 6–9, 2019, doi: 10.1109/ISCAS.2019.8702446.
- [17] H. Li, A. Mehul, J. Le Kernec, S. Z. Gurbuz, and F. Fioranelli, "Sequential Human Gait Classification with Distributed Radar Sensor Fusion," *IEEE Sens. J.*, vol. 21, no. 6, pp. 7590–7603, 2021, doi: 10.1109/JSEN.2020.3046991.
- [18] X. Yang, P. Chen, M. Wang, S. Guo, C. Jia, and G. Cui, "Human motion serialization recognition with through-the-wall radar," *IEEE Access*, vol. 8, pp. 186879–186889, 2020, doi: 10.1109/ACCESS.2020.3029247.
- [19] H. Sadreazami, M. Bolic, and S. Rajan, "Capsfall: Fall detection using ultra-wideband radar and capsule network," *IEEE Access*, vol. 7, pp. 55336–55343, 2019, doi: 10.1109/ACCESS.2019.2907925.
- [20] H. Bin Kwon et al., "Hybrid CNN-LSTM Network for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar," *IEEE Access*, vol. 10, pp. 17556–17564, 2022, doi: 10.1109/ACCESS.2021.3081747.
- [21] H. Sadreazami, M. Bolic, and S. Rajan, "TL-FALL: Contactless Indoor Fall Detection Using Transfer Learning from a Pretrained Model," *Med. Meas. Appl. MeMeA 2019 - Symp. Proc.*, pp. 1–5, 2019, doi: 10.1109/MeMeA.2019.8802154.
- [22] Z. Zhengliang, Y. Degui, Z. Junchao, and T. Feng, "Dataset of human motion status using IR-UWB through-wall radar," *J. Syst. Eng. Electron.*, vol. 32, no. 5, pp. 1083–1096, 2021, doi: 10.23919/JSEE.2021.000093.
- [23] S. Ahmed, D. Wang, J. Park, and S. H. Cho, "UWB-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors," *Sci. Data*, vol. 8, no. 1, pp. 1–9, 2021, doi: 10.1038/s41597-021-00876-0.
- [24] M. Mostafa and S. Chamaani, "Unobtrusive human activity classification based on combined time-range and time-frequency domain signatures using ultrawideband radar," *IET Signal Process.*, no. June, 2021, doi: 10.1049/sil2.12060.

- [25] M. Piriyaajakonkij et al., "SleepPoseNet: Multi-View Learning for Sleep Postural Transition Recognition Using UWB," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 4, pp. 1305–1314, 2021, doi: 10.1109/JBHI.2020.3025900.
- [26] H. Sadreazami, M. Bolic, and S. Rajan, "On the use of ultra wideband radar and stacked lstm-rnn for at home fall detection," *2018 IEEE Life Sci. Conf. LSC 2018*, no. i, pp. 255–258, 2018, doi: 10.1109/LSC.2018.8572048.
- [27] H. Sadreazami, M. Bolic, and S. Rajan, "Fall Detection Using Standoff Radar-Based Sensing and Deep Convolutional Neural Network," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 67, no. 1, pp. 197–201, 2020, doi: 10.1109/TCSII.2019.2904498.
- [28] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, 2019, doi: 10.1109/MSP.2018.2890128.
- [29] B. Erol and M. G. Amin, "Fall motion detection using combined range and doppler features," *Eur. Signal Process. Conf.*, vol. 2016-Novem, pp. 2075–2080, 2016, doi: 10.1109/EUSIPCO.2016.7760614.
- [30] Y. Shao, S. Guo, L. Sun, and W. Chen, "Human motion classification based on range information with deep convolutional neural network," *Proc. - 2017 4th Int. Conf. Inf. Sci. Control Eng. ICISCE 2017*, pp. 1519–1523, 2017, doi: 10.1109/ICISCE.2017.317.
- [31] Y. Li, Z. Peng, R. Pal, and C. Li, "Potential Active Shooter Detection Based on Radar Micro-Doppler and Range-Doppler Analysis Using Artificial Neural Network," *IEEE Sens. J.*, vol. 19, no. 3, pp. 1052–1063, 2019, doi: 10.1109/JSEN.2018.2879223.
- [32] X. Li, Z. Li, F. Fioranelli, S. Yang, O. Romain, and J. Le Kerneec, "Hierarchical radar data analysis for activity and personnel recognition," *Remote Sens.*, vol. 12, no. 14, pp. 1–22, 2020, doi: 10.3390/rs12142237.
- [33] B. Jakanovic, M. Amin, and B. Erol, "Multiple joint-variable domains recognition of human motion," *2017 IEEE Radar Conf. RadarConf 2017*, pp. 0948–0952, 2017, doi: 10.1109/RADAR.2017.7944340.
- [34] K. Bouchard, J. Maitre, C. Bertuglia, and S. Gaboury, "Activity Recognition in Smart Homes using UWB Radars," *Procedia Comput. Sci.*, vol. 170, pp. 10–17, 2020, doi: 10.1016/j.procs.2020.03.004.
- [35] C. Ding et al., "Non-Contact Human Motion Recognition Based on UWB Radar," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 8, no. 2, pp. 306–315, 2018, doi: 10.1109/JETCAS.2018.2797313.
- [36] Y. Wang, J. Zhou, J. Tong, and X. Wu, "UWB-radar-based synchronous motion recognition using time-varying range-Doppler images," *IET Radar, Sonar Navig.*, vol. 13, no. 12, pp. 2131–2139, 2019, doi: 10.1049/iet-rsn.2019.0240.
- [37] J. Park and S. H. Cho, "IR-UWB radar sensor for human gesture recognition by using machine learning," *Proc. - 18th IEEE Int. Conf. High Perform. Comput. Commun. 14th IEEE Int. Conf. Smart City 2nd IEEE Int. Conf. Data Sci. Syst. HPCC/SmartCity/DSS 2016*, pp. 1246–1249, 2017, doi: 10.1109/HPCC-SmartCity-DSS.2016.0176.
- [38] F. Khan, S. Leem, and S. H. Cho, "Hand-based gesture recognition for vehicular applications using IR-UWB radar," *Sensors (Switzerland)*, vol. 17, no. 4, 2017, doi: 10.3390/s17040833.
- [39] A. K. Seifert, A. M. Zoubir, and M. G. Amin, "Radar classification of human gait abnormality based on sum-of-harmonics analysis," *2018 IEEE Radar Conf. RadarConf 2018*, pp. 940–945, 2018, doi: 10.1109/RADAR.2018.8378687.
- [40] G. Wang and Z. Zhu, "Classification of Human Motion Status Using UWB Radar Based on Decision Tree Algorithm," *Int. Conf. Commun. Signal Process. Syst. Springer, Singapore*, 2019., pp. 985–992, 2019, doi: 10.1007/978-981-13-9409-6_115.
- [41] D. Kumar, A. Sarkar, S. R. Kerketta, and D. Ghosh, "Human Activity Classification Based On Breathing Patterns Using IR-UWB Radar," *2019 IEEE 16th India Counc. Int. Conf. INDICON 2019 - Symp. Proc.*, pp. 19–22, 2019, doi: 10.1109/INDICON47234.2019.9029107.
- [42] A. Chowdhury, T. Das, S. Rani, A. Khasnobish, and T. Chakravarty, "Activity recognition using ultra wide band range-time scan," *Eur. Signal Process. Conf.*, vol. 2021-Janua, pp. 1338–1342, 2021, doi: 10.23919/Eusipco47968.2020.9287598.
- [43] W. Chen et al., "Non-Contact Human Activity Classification using DCNN based on UWB Radar," *IEEE MTT-S 2019 Int. Microw. Biomed. Conf. IMBioC 2019 - Proc.*, vol. 1, pp. 1–4, 2019, doi: 10.1109/IMBIOC.2019.8777793.
- [44] H. Du, T. Jin, Y. He, Y. Song, and Y. Dai, "Segmented convolutional gated recurrent neural networks for human activity recognition in ultra-wideband radar," *Neurocomputing*, vol. 396, no. xxxx, pp. 451–464, 2020, doi: 10.1016/j.neucom.2018.11.109.
- [45] S. Ahmed and S. H. Cho, "Hand gesture recognition using an IR-UWB radar with an inception module-based classifier," *Sensors (Switzerland)*, vol. 20, no. 2, 2020, doi: 10.3390/s20020564.
- [46] F. Khan, S. K. Leem, and S. H. Cho, "In-Air Continuous Writing Using UWB Impulse Radar Sensors," *IEEE Access*, vol. 8, pp. 99302–99311, 2020, doi: 10.1109/ACCESS.2020.2994281.
- [47] S. Skaria, D. Huang, A. Al-Hourani, R. J. Evans, and M. Lech, "Deep-Learning for Hand-Gesture Recognition with Simultaneous Thermal and Radar Sensors," *Proc. IEEE Sensors*, vol. 2020-October, pp. 2020–2023, 2020, doi: 10.1109/SENSORS47125.2020.9278683.

Cloud giants: AWS, Azure and GCP

Igor Nevludov, Svitlana Sotnik

*Department of Computer-Integrated Technologies, Automation and Robotics
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine*

Abstract — This work studies cloud giants such as Amazon Web Services, Azure and Google Cloud Platform and then compares them in detail. The work identifies 20 characteristics of their performance. Providers are evaluated for competitiveness, which includes analysis of various aspects and characteristics that are important for organizations and customers using cloud services. A comparative table and diagram based on which, taking into account such key features as security, impact of providers on business processes and IT infrastructure of company showed that each of them has its own features and limitations, which must be known for optimal choice of cloud provider.

Keywords— *Platform, Information Systems, Cloud Technologies, Cloud Providers, Comparison.*

I. INTRODUCTION

Cloud Technology (CT) is nowadays important part of information systems (IS) and technology as it provides infrastructure for data storage and processing [1, 2].

CTs have been widely implemented due to ability to store, process and provide access to data and software solutions via Internet, which is convenient and has become particularly relevant in last three years, which are characterized by fact that most people work remotely. CTs enable both private enterprises and organizations:

- increase scale of their information systems;
- reduce hardware and support costs;
- increase availability of data and services to users.

Modern organizations use cloud platforms to develop and implement their information systems, as well as to provide access to data and applications from any location and device, so topic of paper is and will be relevant.

Informatization is key factor in today's world to improve efficiency and competitiveness of organizations, with IS and cloud technologies interacting and importantly, complementing each other [3, 4].

The topic of this work is cloud giants: Amazon Web Services (AWS), Azure and Google Cloud Platform (GCP), as result of research conducted it is planned to highlight features of their functioning; find out how they compete and what they offer to users; focus of paper will be on security issues and impact of advanced cloud providers on business processes and information technology infrastructure (IT infrastructure).

This topic aroused our interest, because in remote working environment, ability to respond quickly to changes in business environment allows you to manage resources more efficiently and reduce costs. And, as result, it will help you make decision on choosing cloud provider for your business or project.

Therefore, we plan to: review the features of AWS, Azure and GCP; compare them.

II. RELATED WORK

A number of papers have been devoted to study of cloud technologies and providers, where authors reveal importance of choosing right cloud provider and emphasize that AWS, Azure, GCP have their unique features and capabilities that can meet different user needs [5-9].

A review paper on cloud service providers-AWS, Azure, GCP [5]. The authors briefly described: cloud architecture; cloud deployment models; cloud service models and compared AWS, 2023 2nd International Conference on Innovative Solutions in Software Engineering Ivano-Frankivsk, Ukraine, November 29-30, 2023

Azure and GCP on 16 features, as result, they concluded that AWS has higher performance than Azure and Google Cloud, however, security issues are not considered.

The study of cloud data analytics solution is investigated by [6] in which authors selected 20 features to compare Amazon Web Services, Google Cloud Platform, Microsoft Azure, authors, make subjective conclusion that only reason why customers should choose GCP cloud provider is because it has best pricing structure for services.

Another new study [7] focusing on performance evaluation when deploying dockerized web application on AWS, GCP and Azure. The authors in this study tested JMeter, SysBench and Apache Benchmark in Docker Container environment deployed on AWS, GCP and Azure, however, nothing is said about limitations for testing.

In [8], performance of existing MQTT brokers operating in AWS, Microsoft Azure and Google Cloud Platform cloud services is analyzed. The authors emphasize on IoT implementation which requires MQTT protocol with high communication performance, fast response and small size.

Security issues and problems in cloud computing are analyzed in [9]. The authors reviewed existing types of services.

III. METHODOLOGY

This research will conduct a comparative study of AWS, Microsoft Azure, and Google Cloud Platform cloud providers to evaluate their features, capabilities, and performance. This type of research will help users and organizations make more informed decisions about choosing cloud provider based on their specific needs and requirements.

This study will address impact of advanced cloud providers on business processes and IT infrastructure and as result, determine what are differences and similarities between AWS, Microsoft Azure and GCP cloud providers in context of their performance, functionality and their security.

First, quick word about AWS. AWS offers more than 200 services including computing, storage, databases, analytics, artificial intelligence, Internet of Things (IoT), and more. AWS has data centers in different parts of world [10, 11].

Users can increase or decrease capacity in real time, allowing for efficient resource utilization and cost reduction. AWS provides wide range of tools and services to ensure data security and privacy. They comply with standards and regulatory requirements, making them suitable for large number of industries.

Azure and GCP also offer more than 200 services and services, just like AWS. Azure is integrated with wide range of Microsoft products and services, including Windows Server, Active Directory, SQL Server and many others, and this makes it easy to migrate existing applications and infrastructure to cloud.

Azure supports hybrid cloud model, allowing organizations to combine cloud resources with on-premises infrastructure. Microsoft is actively investing in Azure security, providing tools and certifications to protect customer data and applications [13, 14].

Rounding out Google Cloud Platform. GCP has powerful resources and services that enable development and deployment of machine learning applications and models, including tools like TensorFlow, AI Platform and BigQuery ML.

GCP offers BigQuery, high-performance and scalable real-time data analytics system that enables analysis and visualization of data large volumes. GCP also, like first two providers, provides high level of security for its customers [14, 15].




Let's highlight characteristics by which comparison will be made (Table 1).

The main criterion is security of these providers and we will emphasize assessment of each provider impact on business processes and IT infrastructure.

In Table 1: 1* Containerization is virtualization method that allows you to package and run applications and their dependencies in isolated containers [16]; 2* Server less solutions is architectural approach to developing and deploying applications where developers do not have to

worry about managing infrastructure, including virtual servers (VMs) or containers. Instead, developers focus on writing application code and its business logic, and cloud provider automatically manages infrastructure, scaling, and execution of code [17, 18].

Table 1 – Comparison of cloud providers features

No	Features	AWS 	Azure 	GCP 
1	Market leadership	Largest provider, market leader	The second largest provider in market	The third largest provider in market
2	Scalability	26 regions, 81 available zones	60 regions, 140 available zones	24 regions, 76 available zones
3	Service Level Agreement (SLA)	Subject to change depending on services		
4	Integration with other services	Wide range of integrations and services	Deep integration with Microsoft	Many integrations and services
5	Virtual machines	EC2	Virtual Machines	Compute Engine
6	Network services	VPC, Direct Connect	Virtual Network, ExpressRoute	VPC, Interconnect
7	Databases	RDS, DynamoDB	Azure SQL Database, Cosmos DB	Cloud SQL, Firestore
8	Analytics and machine learning	Amazon EMR, SageMaker	Azure Machine Learning	AI Platform, BigQuery
9	Data storage	S3, EBS, Glacier	Azure Blob Storage, Azure Disk	Cloud Storage, Persistent Disk
10	Identification and access control	IAM, Cognito	Azure Active Directory	Identity and Access Management
11	Containerization (1*)	Amazon ECS, EKS	Azure Kubernetes Service (AKS)	Google Kubernetes Engine (GKE)
12	Serverless solutions (2*)	AWS Lambda	Azure Functions	Cloud Functions
13	Security Management Center	AWS Identity and Access Management (IAM) provides flexible access control.	Azure Active Directory (Azure AD) for access control and authorization. Provides flexible access control.	Google Cloud Identity and Access Management (IAM) provides flexible access control.
14	Data access	AWS Identity and Access Management (IAM) allows you to customize access policies.	Azure Active Directory (Azure AD) and Azure RBAC for access control.	GCP Cloud Identity & Access Management provides granular access control.
15	Physical security	Ensures physical security of its data centers and networks.		
16	Data protection	AWS also provides means to encrypt	Azure provides data encryption	Google Cloud offers wide range of tools

		data in different layers.	and key management tools.	for encrypting data at rest and in motion.
17	Monitoring and threat detection	AWS offers Amazon GuardDuty for threat detection and security monitoring.	Azure Security Center for threat monitoring and detection.	Google Cloud offers Chronicle for threat detection and analysis.
18	Flexibility	Highly flexible to respond quickly to changing business needs.	Mid-level flexibility, more integrated with Microsoft solutions.	Good flexibility, especially when combined with Kubernetes and other tools.
19	Deployment speed	Rapid deployment of resources and services, enabling faster launch of new products and projects.	Deployment may take longer, but is integrated with other Microsoft products.	Relatively fast deployment, especially using templates and automation.
20	Cost and economic efficiency	AWS provides wide range of pricing models and payment options that can help optimize costs.	Azure also offers flexible pricing options, but is integrated with other Microsoft solutions.	GCP offers similar pricing options, with emphasis on simplicity and transparency in pricing.

The choice of cloud provider depends on project developers specific requirements, budget and preferences.

The selected characteristics will help determine which cloud provider best meets the needs of particular type of business and IT infrastructure. It is also important to consider specifics of organization, project timelines and budget constraints.

IV. RESULTS

The results of research conducted will help determine characteristics of each these giants, and hallmark of this study is to determine impact of these cloud providers on business processes and IT infrastructure while taking into account security issues addressed. As result, choosing best partner for cloud services.

So, having analyzed cloud providers according to selected criteria (Figure 1).

Figure 1 proposes diagram in which providers have been evaluated for competitiveness, which includes analyzing various aspects and characteristics. The cloud providers (AWS, Azure and GCP) are compared on a scale: 1 – best option, 2 – average option and 3 – worst option with emphasis on security of these providers and evaluating impact each of them on business processes and IT infrastructure.

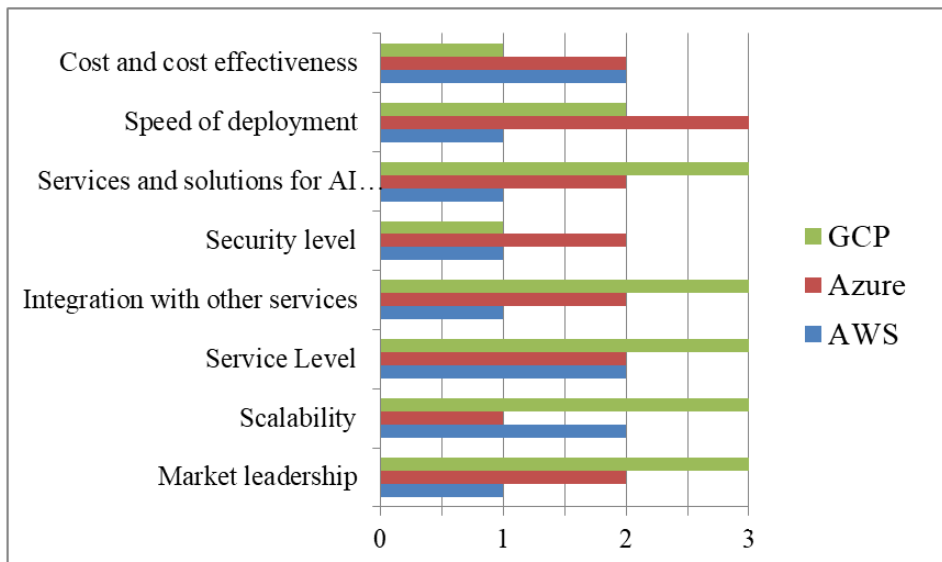


Figure 1 – Diagram comparing AWS, Azure, and GCP

Based on research conducted, following conclusions can be drawn:

- AWS is leader with high security, scalability, and wide range of integrated services. AWS is rated by us as best option;
- Azure provides medium level of services and security, with deep integration into Microsoft ecosystem – this is middle option;
- GCP provides good level of security and fast deployment, but is rated as worst option due to lower scalability and service levels.

V. DISCUSSION

Taking into account results of research conducted to determine impact of advanced cloud providers such as AWS, Azure, GCP on business processes, information technology infrastructure and based on analysis conducted taking into account performance, functionality and security of these providers, several key conclusions can be drawn:

- AWS is leading the way, as this result is evidenced by its high level of security, flexibility and extensive set of integrated services. AWS gives companies ability to build and grow complex information infrastructures with high degree of reliability and security. Companies that are scalability-oriented and ready to invest in cloud solutions can find optimal partner in AWS;
- mid-tier – Azure, as Microsoft Azure provides mid-tier level of services and security. It stands out for its deep integration into Microsoft ecosystem, making it attractive choice for companies that already use Microsoft products or wish to integrate them into cloud environment. However, Azure can be less flexible than AWS, which can require additional effort in planning and resource management;
- relatively "low tier" – GCP, although this provider provides good level of security and fast application deployment. However, GCP's ranking as worst option is due to its more limited scalability and service levels. This provider may be attractive to innovation-oriented companies willing to utilize Google's advanced technologies, but it may require additional effort to ensure scalability as business grows.

In any case, it is important to consider specifics of each particular organization, project timelines and budget constraints.

VI. FUTURE RESEARCH

Future research can investigate timing of real-world projects and service level definitions.

VII. CONCLUSION

As result of research conducted on cloud giants such as Amazon Web Services, Azure and Google Cloud Platform and then comparing them in detail, 20 characteristics of their performance are highlighted. A competitive evaluation of providers has been done, which includes analysis of various aspects and characteristics that are important to organizations and customers using cloud services. A comparative table and diagram based on which, taking into account such key features as security, impact of providers on business processes and IT infrastructure of company showed that each of them has its own features and limitations, is given.

VIII. DISCLOSURES

We declare that there are no conflicts of interest associated with this work.

REFERENCES

- [1] V. Lyashenko, "Development Features Web-Applications," *International Journal of Academic and Applied Research (IJAAR)*, 2023, vol. 7, issue 1, pp. 79-85.
- [2] Z. Deineko, "Features of Database Types," *International Journal of Engineering and Information Systems (IJEAIS)*, 2021, vol. 5, issue 10, pp. 73-80.
- [3] S. Sotnik, "Key Directions for Development of Modern Expert System," *International Journal of Engineering and Information Systems (IJEAIS)*, 2022, vol. 6, issue 5, pp. 4-10.
- [4] I. A. Borysenko, "Chat gpt features in data search," *The 9th International scientific and practical conference "Scientific progress: innovations, achievements and prospects" (May 29-31, 2023) MDPC Publishing*, 2023, pp. 139-143.
- [5] A. Yevge, "Review Paper on Cloud Service Provider–AWS, Azure, GCP," *EasyChair*, 2022. – №. 7440.
- [6] N. M. Gupta, "AWS VS AZURE VS GCP: LEADERS OF THE CLOUD RACE," *Int. Res. J. Mod. Eng. Technol. Sci*, 2022, vol. 4, issue 7, pp. 3243-3250.
- [7] A. P. Kurniawan, "Performance Evaluation for Deploying Dockerized Web Application on AWS, GCP, and Azure," *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*, 2023, pp. 346-350.
- [8] A. S. S. Ansyah, "MQTT Broker Performance Comparison between AWS, Microsoft Azure and Google Cloud Platform," *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)*, 2023, pp. 1-6.
- [9] R. Rani, "Security Issues and Challenges in Cloud Computing," *Multidisciplinary Approach in Arts, Science & Commerce (Volume-3)*, 2023, pp. 82.
- [10] T. Singh, "The effect of Amazon Web Services (AWS) on Cloud-Computing," *International Journal of Research and and Technology (IJERT)*, 2021, vol. 10, issue 11, pp. 2278-0181.
- [11] P. Mishra, "Advanced AWS Services," *Cloud Computing with AWS: Everything You Need to Know to be an AWS Cloud Practitioner*, 2023, pp. 247-277.
- [12] S. J. Arévalo-Cordero, "Data traffic management in hybrid cloud composed of Openstack and Azure," *2019 IEEE Colombian Conference on Communications and Computing (COLCOM)*, 2019, pp. 1-6.
- [13] N. Bhardwaj, "Case Study of Azure and Azure Security Practices," *Machine Learning Techniques and Analytics for Cloud Security*, 2021, pp. 339-355.
- [14] A. Roy, "Study on Google Cloud Platform (GCP) and Its Security," *Machine Learning Techniques and Analytics for Cloud Security*, 2021, pp. 313-338.
- [15] E. Bisong, "An overview of google cloud platform services," *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 2019, pp. 7-10.
- [16] G. I. Radchenko, "Comparative analysis of virtualization methods in big data processing," *Supercomputing Frontiers and Innovations*, 2019, vol. 6, no. 1, pp. 48-79.
- [17] M. Kumar, "Serverless architectures review, future trend and the solutions to open problems," *American Journal of Software Engineering*, 2019, vol. 6, no. 1, pp. 1-10.
- [18] L. Jiang, "Overview of serverless architecture research," *Journal of Physics: Conference Series*, 2020, Ser. 1453, №. 1, pp. 1-8.

Адаптивний захист інформації на мобільному пристрої

Євген Бровченко, Валерій Самарай, Володимир Павленко

*Інститут комп'ютерних технологій
Відкритий міжнародний університет розвитку людини «Україна»
Київ, Україна*

Анотація—Дана робота присвячена проблемі використання мобільного пристрою в умовах активного впливу та захист неструктурованої інформації. Увага приділена використанню мобільного пристрою як частини інформаційної системи та можливі заходи щодо захисту інформації від протиправних дій. Була проведена робота по вивченню проблеми в реальних умовах використання та взаємодії. Пропонується модель алгоритму адаптивного захисту інформації. В основу моделі покладена ідея динамічної адаптації системи виходячи з поточного стану та при умові активного впливу.

Ключові слова—*Мобільний пристрій, Захист інформації, Неструктурована інформація, Кібербезпека.*

I. ВСТУП

Інформаційні технології та обчислювальні системи з кожним днем все більше стають невід'ємною частиною життя людини, незамінними помічниками а іноді й джерелом проблем. Використання сучасних інформаційних технологій у багатьох сферах життя вимагає від інженерів та розробників нових рішень, необхідно закласти міцний фундамент для подальшого розвитку та можливість ефективного удосконалення чи підтримки. Розвиток технологій відбувається зі швидкістю блискавки та має великий вплив на наше суспільство, кожний день ми маємо можливість спостерігати нові технічні та інженерні винаходи. Використання штучного інтелекту у повсякденному житті вже не є чимось фантастичним. Причому це може бути як певна інтеграція у робочий процес – автоматизація повсякденної рутини, так і частина дозвілля – спілкування та розваги, тощо. Весь комплекс технологій та систем поліпшують та спрощують виконання повсякденних задач, відкривають нові можливості та прибирають кордони. Важко уявити собі сучасну людину без мобільного телефону чи планшета, що має майже безперервний та безлімітний доступом до Інтернету чи всевітньо глобальної мережі. Мобільний пристрій посідає своє важливе місце у житті сучасної людини. В подальшому інтеграція людини та мобільного пристрою матиме глибший характер та непередбачуваність. Тому треба розглядати мобільний пристрій як не від'ємну складову сучасних інформаційних систем. Сьогодні мобільний пристрій має певний сакральний сенс, він може містити багато різноманітної інформації чи даних. Природа та важливість цієї інформації може бути різною, але захист має пріоритетне значення. А враховуючи динаміку розвитку та невизначеність деяких процесів, перед інженерами та розробниками стоїть важке та важливе завдання по захисту інформаційних систем та забезпеченню надійної роботи. До певної міри мобільний пристрій можна вважати найбільш проблематичною складовою, попередні рішення не завжди можуть забезпечити бажаний результат. Цей пристрій має ряд переваг та особливостей – власне це й роблять його уразливим перед потенційними зловмисниками та й дії користувача можуть спричинити ряд проблем. Дана робота присвячена проблематиці використання мобільних пристроїв та захисту неструктурованої інформації. Захист неструктурованої інформації на мобільному пристрої є важливим завданням з точки зору кібербезпеки. Неструктурована інформація може бути у вигляді текстових документів, електронних листів, фотографій, відео, записів голосу та іншого

контенту, який зберігається на мобільних пристроях. До основних проблем можна віднести: втрата або крадіжка мобільного пристрою; вразливості операційної системи та програмного забезпечення; недостатній рівень шифрування та недостатній рівень складності паролів; нехтування кібергігієною. Існуючі інженерно-архітектурні рішення не завжди виконують поставлену задачу та можуть мати проблеми з інтеграцією. Не слід забувати й про економічну складову. Аналіз прикладної області та результати дослідження говорять про те, що захист інформації в цілому є важливим та є попит на швидкі та ефективні рішення щодо безпеки інформаційних систем в цілому та мобільного пристрою зокрема. Треба зазначити що питання безпеки є комплексним: починаючи від етапу створення та контроль на етапі експлуатації, також дуже важливим є інформаційна обізнаність та відповідальність кінцевого користувача. Кібербезпека виходить на принципово новий рівень. Лише комплексний підхід може забезпечити ефективне рішення. Динамічна адаптації системи безпеки має розглядатись як основа стабільної інформаційної системи.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Було проведено глибоке теоретичне та практичне досліджене предметної області, розглянуто особливості використання мобільних пристроїв як індивідуального використання так і корпоративного рівня. Це дало змогу визначити загальні тенденції і методи, які використовуються в області адаптивного захисту інформації на мобільних пристроях. Основу складає біометричний доступ. Біометричні дані, такі як відбитки пальців, розпізнавання обличчя та інші біометричні методи, широко використовуються для забезпечення доступу до мобільних пристроїв. Дослідження в області біометрії постійно вдосконалюють ці технології для забезпечення більшої точності та безпеки. Звісно не треба забувати про шифрування, оскільки це важливий аспект захисту інформації на мобільних пристроях. Ця область розвивається, алгоритми шифрування та методи зберігання ключів вдосконалюються повсякденно. Дещо новим для мобільного використання стали віртуальні приватні мережі (VPN). Використання VPN дозволяє захищати інформацію від несанкціонованого доступу та забезпечує приватність користувача під час використання глобальної мережі, але це рішення не завжди ефективно використовується кінцевим користувачем. Наступним кроком можна вважати захист від вірусів і шкідливих програм. Сучасні вчені та спеціалісти в області кібербезпеки постійно вивчають нові загрози, що стосуються мобільних пристроїв, та розробляють програми та методи для виявлення та видалення шкідливих програм. Поряд з ефективністю біометрії можна поставити мультифакторну аутентифікацію. Цей напрям аутентифікації продовжує розвивати ефективні методи, які вимагають від користувача подання декількох факторів, таких як пароль, біометричні дані та одноразові коди. Різноманіття рішень та рівень інтеграції інколи ускладнює сприйняття кінцевого користувача, а винахідливість зловмисників може спричинити великі втрати. Як результат аналізу було виявлено, що адаптивний захист має знайти ефективне рішення на поставлену проблему захисту інформації на мобільному пристрої. Ця концепція полягає в тому, щоб застосовувати захист інформації в залежності від динамічного контексту користувача та середовища. Наприклад, захист може бути посилено в залежності від місця розташування користувача, часу доби, стану пристрою тощо.

III. МЕТОДИ

Обчислювальна система – це сукупність апаратних і програмних компонентів, які взаємодіють для вирішення обчислювальних завдань. Обчислювальні системи можуть включати різноманітні пристрої, такі як комп'ютери, мобільні телефони, планшети, сервери, ноутбуки, суперкомп'ютери. Слід розуміти, що перелік цих складових може змінюватись, причому деякі пристрої можуть зникати, а про деякі ми сьогодні навіть і не здогадуємось. Обчислювальні системи можуть бути використані для різноманітних завдань, таких як

обробка даних, створення документів, візуалізація інформації, аналіз даних, ігри та інше. Вони можуть бути використані в різних галузях, таких як наука, бізнес, медицина, техніка, транспорт, тощо. Рівень інтеграції обчислювальних систем у повсякденне життя з часом буде лише збільшуватись та набувати нових рис.

Мобільний пристрій – це портативний електронний пристрій, який має можливість підключення до мережі зв'язку та доступу до різноманітних сервісів та додатків. У сучасному світі мобільні пристрої стали невід'ємною частиною життя людей, і це змінило спосіб, яким ми спілкуємося, взаємодіємо, працюємо та розважаємося. Однією з основних функцій мобільного пристрою є забезпечення зв'язку, що дозволяє користувачам здійснювати дзвінки та відео дзвінки, надсилати та отримувати повідомлення, електронну пошту та інші форми комунікації. Мобільні пристрої також забезпечують доступ до Інтернету, що дає користувачам можливість шукати інформацію, переглядати відео, грати в ігри, слухати музику та виконувати інші дії. Мобільні пристрої також стали необхідним інструментом для роботи та навчання, дозволяючи користувачам працювати з електронною поштою, текстовими документами, електронними таблицями та іншими програмами. Вони можуть бути використані для спостереження за здоров'ям та фітнесом, ведення щоденника, планування та організації розкладу та інших аспектів життя. Однією з основних переваг мобільних пристроїв є їх портативність та доступність, що дозволяє людям бути зв'язаними та отримувати інформацію в будь-якому місці та в будь-який час. У цілому, мобільні пристрої стали невід'ємною частиною життя сучасної людини, і вони продовжують еволюціонувати та вдосконалюватися. Завдяки їхній функціональності та зручності використання, мобільні пристрої дозволяють людям бути більш продуктивними та організованими в їхньому житті та роботі. Можна відділити такі основні типи мобільних пристроїв: смартфон, планшет, ноутбук, ультрабук, різноманітні гаджети, які можуть бути одягнуті на тіло, наприклад, смарт-годинники, фітнес-браслет. Існують основні операційні системи: Android, iOS, Windows/Mobile, BlackBerry, KaiOS. Різноманіття мобільних пристроїв з одного боку дає можливість обрати саме те, що треба користувачу, з іншого боку ускладнює роботу проектувальників та розробників.

Особливу увагу слід приділити Android Enterprise. Android Enterprise (раніше відомий як Android for Work) – це програма та набір функцій, розроблених компанією Google для підприємств, які використовують мобільні пристрої на базі операційної системи Android. Android Enterprise надає розширені можливості керування і захисту даних, що дозволяє організаціям легко і безпечно інтегрувати та керувати пристроями Android в корпоративному середовищі.

Деякі ключові функції та можливості Android Enterprise включають:

1. Робочі профілі: дозволяє розділити особисті дані і програми від корпоративних на пристрої з Android. Це дозволяє зберігати особисті дані окремо від корпоративних даних і додатків.
2. Управління пристроями: забезпечує централізоване управління пристроями, що дозволяє адміністраторам віддалено керувати налаштуваннями, політиками безпеки, встановлювати додатки та проводити оновлення.
3. Політики безпеки: дозволяє встановлювати політики безпеки на пристроях, такі як вимоги до паролів, шифрування даних, обмеження функцій пристрою та інші заходи безпеки.
4. Магазин додатків для підприємств: надає можливість створювати власний магазин додатків для підприємств, де адміністратори можуть контролювати доступні додатки та встановлювати їх на пристрої.

5. Інтеграція з обліковими записами Google Workspace: компанії, які використовують Google Workspace (раніше G Suite), можуть легко інтегрувати свої корпоративні облікові записи з Android Enterprise і керувати доступом до облікових записів та сервісів.

6. Захист даних: можна використовувати різні методи захисту даних, такі як шифрування пристроїв, вимоги до паролів, обмеження прав доступу та інші політики безпеки, щоб забезпечити конфіденційність і цілісність даних.

7. Інтеграція з іншими системами: може інтегруватися з різними корпоративними системами, такими як системи керування ідентифікацією (IAM), системи електронної пошти та календаря, хмарні сервіси для зберігання даних та багато іншого.

8. Завдання та автоматизація: створювання різних завдань та автоматизування процесів на пристроях, що дозволяє підвищити продуктивність та ефективність користувачів.

Для захисту мобільного пристрою та інформації досить ефективною є використання біометричної ідентифікації. Біометрія – це використання фізичних або поведінкових характеристик людини для ідентифікації, аутентифікації та авторизації. На сучасних мобільних пристроях все більше використовуються біометричні методи для захисту неструктурованої інформації, що зберігається на них. Використання біометрії дозволяє підвищити безпеку і зручність доступу до цієї інформації.

Технології біометричної ідентифікації на мобільних пристроях:

- Сканування відбитків пальців: сканер вбудований в сенсорний екран або на задній панелі пристрою. Відбиток пальця порівнюється з збереженими біометричними даними, що дозволяє впізнати користувача.
- Розпізнавання обличчя: камера пристрою використовується для захоплення обличчя користувача, яке порівнюється з попередньо збереженими біометричними даними.
- Розпізнавання відбитків очей: сітківка може бути використані для ідентифікації користувача. Спеціальні камери або сенсори захоплюють деталі очей і порівнюють їх з шаблонами біометричних даних.
- Розпізнавання голосу: вбудований мікрофон записує голос користувача, який порівнюється з попередньо збереженими голосовими шаблонами.

Переваги використання біометрії на мобільних пристроях:

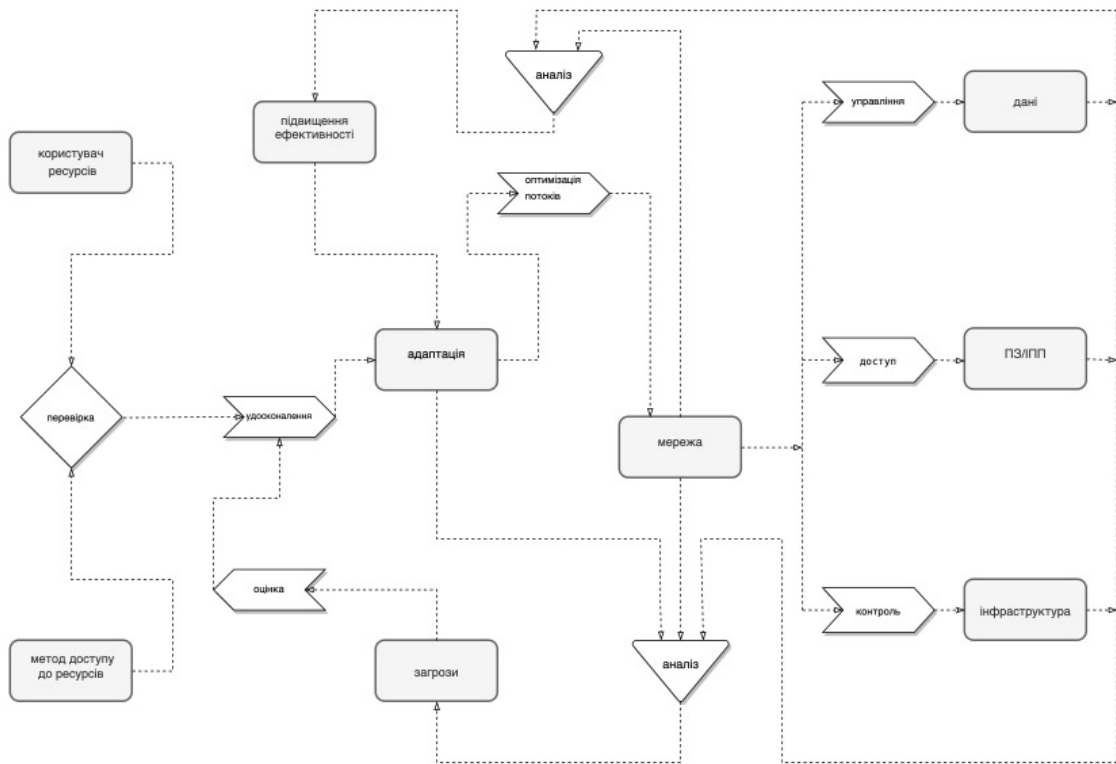
- Вищий рівень безпеки. Біометричні дані є унікальними для кожної особи і майже неможливо підробити або підмінити. Це дозволяє забезпечити високий рівень ідентифікації та аутентифікації користувача.
- Зручність використання. Використання біометрії на мобільних пристроях забезпечує зручний та швидкий доступ до інформації.
- Ізольованість зберігання. Біометричні дані зазвичай зберігаються в зашифрованому вигляді на самому пристрої. Це зменшує ризик їх витоку або несанкціонованого доступу до них.
- Можливості багатофакторної аутентифікації. Біометричні методи можуть бути поєднані з іншими методами аутентифікації, такими як пін-коди або паролі, для створення більш надійного і багатоетапного процесу ідентифікації.
- Необхідність фізичного присутності. Використання біометрії вимагає фізичної присутності особи для ідентифікації. Це ускладнює можливість несанкціонованого доступу до пристрою або даних.

Наступним кроком можна вважати захист інформації на рівні операційної системи чи мобільно додатку. MAM (Mobile Application Management) є скороченням від англійського терміну "Mobile Application Management". Він відноситься до підходу до управління мобільними додатками в організації. MAM використовується для керування та контролю мобільними додатками, які встановлюються на пристрої користувача. Це може бути застосунок для внутрішнього використання в організації або публічний додаток, який користувачі завантажують з магазинів додатків. Основна мета MAM полягає в управлінні додатками на рівні самого додатку, незалежно від пристрою. Вона надає можливості забезпечити безпеку, контроль та конфігурацію додатків, а також здійснювати моніторинг, оновлення та віддалене управління додатками на мобільних пристроях. Деякі можливості MAM можуть включати: можливість встановлення та виконання політик, що стосуються використання додатків, таких як обмеження функцій або доступу до даних; здатність забезпечувати безпеку додатків шляхом використання шифрування, контролю доступу та ідентифікації користувачів; можливість розповсюдження додатків на пристрої користувачів через магазини додатків або корпоративні канали, а також віддалене оновлення додатків; здатність збирати дані про використання додатків, продуктивність та поведінку користувачів для аналізу та прийняття відповідних рішень. MDM (Mobile Device Management) є скороченням від англійського терміну "Mobile Device Management". Це підхід до управління мобільними пристроями в організації. MDM використовується для централізованого керування, контролю та захисту мобільних пристроїв, таких як смартфони та планшети, що використовуються в організації. Він надає можливості забезпечення безпеки, налаштування, моніторингу та віддаленого управління мобільними пристроями з одного централізованого інтерфейсу. Основна мета MDM полягає в забезпеченні безпеки, конфігурації та керування мобільними пристроями в організації. Деякі можливості MDM включають: можливість реєстрації мобільних пристроїв в системі MDM та їх прив'язка до конкретних користувачів або груп; здатність встановлення та забезпечення політики безпеки на мобільних пристроях, включаючи налаштування паролів, шифрування даних, віддалене видалення даних тощо; можливість віддалено налаштувати пристрої та встановлювати додатки, контролювати доступ до магазинів додатків та обмежувати встановлення небезпечних або небажаних додатків; здатність збирати дані про стан та використання мобільних пристроїв, включаючи інформацію про батарею, сигнал мережі, використання пам'яті.

Не слід забувати й про користувача інформаційних систем. Користувач подекуди може бути найслабкішим та вразливішим елементом. Обізнаність користувача щодо безпеки при користуванні мобільним пристроєм є важливим аспектом для забезпечення захисту особистих даних і запобігання кіберзлочинності. Всі ці моменти слід забезпечувати, контролювати та поліпшувати. Можна виділити основні аспекти: захист паролів: користувачі повинні встановлювати міцні паролі або використовувати біометричні методи аутентифікації для захисту своїх пристроїв; важливо використовувати унікальні та складні паролі і не ділитися ними з іншими особами; оновлення програмного забезпечення регулярно, оновлення операційної системи та програмного забезпечення на пристрої є важливим, оскільки виробники випускають оновлення та виправлення для заповнення потенційних слабких місць; встановлення довірених додатків, користувачі повинні завантажувати додатки тільки з офіційних магазинів додатків, таких як Google Play або App Store, і перевіряти рейтинги та відгуки інших користувачів перед встановленням; уважність при відкритті посилань та відправці особистої інформації, користувачі повинні бути обережними при натисканні на посилання в недовіреному повідомленні або електронній пошті та уникати надсилання особистих чутливих даних через ненадійні канали комунікації; використання антивірусного програмного забезпечення, встановлення добре відомого антивірусного програмного забезпечення на мобільний пристрій може допомогти виявити і запобігти інфікуванню шкідливим програмним забезпеченням.

IV.РЕЗУЛЬТАТИ

В результаті проведеного аналізу та дослідження убуло запропоновано модель алгоритму адаптивного захисту інформації.



Мал. 1

Основні компоненти:

- Користувач
- Метод доступу до ресурсів
- Мережа
- Інформація
- Програмне забезпечення
- Інфраструктура
- Загрози

На основі моніторингу, аналізу та оптимізації приймається рішення про певний набір дій та протидій що має призвести до підвищення ефективності шляхом адаптація до конкретних умов. Управління проходить в автоматичному, напівавтоматичному та ручному режимі.

V.ОБГОВОРЕННЯ

Запропонований підхід дає можливість швидко реагувати на будь-які зміни, що відбуваються в середині системи, а також на впливи зовні. Адаптація відбувається з урахуванням поточних потреб та можливостей кінцевої інфраструктури. Компоненти та взаємодія між ними відбувається з урахуванням твердження, що атака чи витік вже відбувся. Це дозволяє планувати певну автономність та незалежність, створювати можливі сценарії.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Планується створення дослідного стенду та подальша оптимізація моделі.

VII. ВИСНОВКИ

Майбутнє захисту інформації на мобільних пристроях буде засноване на поєднанні технологічних нововведень та підвищенні обізнаності користувачів з питань кібербезпеки. Деякі напрямки розвитку включають застосування штучного інтелекту (AI) та машинного навчання: AI може допомогти в розробці розширених механізмів захисту, що автоматично виявляють аномальні або підозрілі дії та блокують потенційні загрози; біометрична автентифікація, покращення технологій біометричної автентифікації, таких як розпізнавання обличчя, відбитків пальців, можуть забезпечити більш надійний захист неструктурованої інформації; застосування блокчейн-технологій: блокчейн може відігравати важливу роль у забезпеченні безпеки інформації, оскільки він дозволяє зберігати дані розподілено та захищено від змін; розвиток квантово-стійких алгоритмів шифрування: з розвитком квантових комп'ютерів створюється потреба в нових алгоритмах шифрування, які можуть витримати потужність квантового злому; приватність за замовчуванням (Privacy by Design): проектування мобільних пристроїв та застосунків з врахуванням приватності користувачів сприятиме захисту неструктурованої інформації на мобільних пристроях; освіта та обізнаність користувачів: підвищення обізнаності користувачів щодо кібербезпеки та навчання їх ефективним методам захисту інформації на мобільних пристроях допоможе зменшити ризик виникнення інцидентів, пов'язаних з безпекою; розвиток законодавства та нормативної бази, вдосконалення законодавства та нормативної бази у галузі кібербезпеки сприятиме створенню надійного середовища для захисту неструктурованої інформації. Захист неструктурованої інформації на мобільних пристроях спиратиметься на комбінації технологічного прогресу, навчання користувачів та співпраці між різними сторонами, такими як розробники, постачальники послуг та законодавці. Забезпечення безпеки та приватності даних в цифровому світі залишається пріоритетом, оскільки кількість мобільних пристроїв та їх використання продовжують зростати.

Сучасний користувач мобільних пристроїв має певні вимоги, що спарюються на ергономіку та рівень інформаційної обізнаності. Розвиток галузі в цілому вимагає швидких, надійних та ефективних рішень, що потрапляють не лише в економічну площину, а й площину кібербезпеки. Використання мобільних пристроїв може адаптуватися до швидких сучасних потреб безпеки організації та може бути легко налаштований відповідно до конкретних вимог. Мобільний пристрій є зручним і доступним для багатьох користувачів, він зазвичай знаходиться поруч з користувачем і може використовуватися як додатковий елемент безпеки.

ЛІТЕРАТУРА

- [1] Choi, J. S., & Kim, H. S. (2019). An Efficient Encryption Method for Securing Unstructured Data on Mobile Devices. *IEEE Access*, 7, 124616-124626.
- [2] Kremer, Helmut, ed. *Mobile Computing: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2018.
- [3] Liu, Z., Huang, L., Gao, Y., Wang, Z., & Yang, X. (2017). Research on cloud computing security and privacy protection model. *Cluster Computing*, 20(3), 1893-1901.
- [4] Madden J., Madden B. *Enterprise Mobility Management: Everything you need to know about MDM, MAM, and BYOD*. – Jack Madden, 2013. 176 p.
- [5] Gámez, N., García-Sánchez, P., & Pérez, J. (2020). User authentication using mobile devices: A review. *Information Fusion*, 55, 153-170.
- [6] Kang, Y., Zhang, N., Zhu, X., & Huang, Y. (2021). A novel authentication mechanism based on dynamic cognitive keystroke on mobile devices. *IEEE Access*, 9, 32098-32107.
- [7] Yavuz, A. G., Akgün, M. A mobile-based multi-factor authentication system using biometric and location information. *International Journal of Computer Science and Network Security*, 17(5), 181-187, (2017).
- [8] Бровченко, С. М., Самарай, В. П., Даценко, І. П., Павленко, В. І., Серета А.В. (2023). Захист неструктурованої інформації на мобільному пристрої. *Інфокомунікаційні та комп'ютерні технології*, 1(05) 2023, С. 194-200.

Entropy, Gaussian Distribution and Fractional Processes

Anatoliy Malyarenko*, Yuliya Mishura,
Kostiantyn Ralchenko and Yevheniia Anastasiia Rudyk

*Division of Mathematics and Physics, Mälardalen University,
Västerås, Sweden**

*Department of Probability Theory, Statistics and Actuarial Mathematics,
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine*

Abstract—We consider five types of entropies for Gaussian distribution: Shannon, Rényi, generalized Rényi, Tsallis and Sharma–Mittal entropy, establishing their interrelations and their properties as the functions of parameters. Then, we consider fractional Gaussian processes and compare the entropies of one-dimensional distributions of these processes. In other words, we compare the information connected to these processes.

Keywords—Shannon entropy; Rényi entropy; Tsallis entropy; Sharma–Mittal entropy; normal distribution; fractional Brownian motion; subfractional Brownian motion; bifractional Brownian motion; multifractional Brownian motion; tempered fractional Brownian motion.

I. INTRODUCTION

The concept of entropy for a random variable was introduced by Shannon [1] to characterize the irreducible complexity inherent in a specific form of randomness. The concept of entropy as a measure of the chaos of a dynamical system has been known for a long time, and this concept is used in numerous applications, starting with the information theory, machine learning, information security, cryptography [2] and continuing with environmental sciences, genetics [3], market analysis [4], climate analysis [5], social network analysis [6], chemical reactions, medical measurements [7,8]. Entropy plays a pivotal role in signal processing and network traffic analysis. It is employed in the development of algorithms for detecting DDoS attacks [9], data compression, decision tree construction [10]. Fractional processes serve as essential models for capturing long-range dependence and self-similarity in diverse data types. Entropy plays a crucial role in quantifying the complexity and information content of signals generated by fractional processes, which proves invaluable for tasks like prediction, risk assessment and anomaly detection.

From a mathematical point of view, the entropy of a probability distribution is expressed in terms of its density, given the specified density, and this entropy is not difficult to calculate for specific distributions. Note, however, that there are many different approaches to determining the entropy of a probability distribution, starting with Shannon entropy, and then this concept was successively complicated and generalized by adding new parameters (Rényi, generalized Rényi, Tsallis[11,12], Sharma–Mittal entropies [13, 14]). The various definitions of entropy share several basic properties postulated by Alfred Rényi [15]. All of the indicated entropies can be successfully calculated (and they have already been calculated, for example, in [16]) in the case of a Gaussian distribution, which is the subject of this paper. However, in the presence of additional parameters of the entropy itself (not the distribution), the question immediately arises about the behavior of entropy as a function of the parameter. It is a well-known fact that the Rényi entropy as a function of the parameter decreases. However, its convexity is not a universal property and, in general,

depends on the distribution [17]. Therefore, if we concentrate on Gaussian distribution, we need to investigate the properties of the introduced entropies in as much detail as possible.

II. METHODOLOGY

Investigation of the properties of the various entropies for the centered Gaussian distribution with respect to the parameters was generally done by applying calculus methods.

The entropy of vector fractional Gaussian noise is calculated using the formulas given in the book [18]. However, firstly, these calculations are based on the fact that fractional Gaussian noise is a stationary process, and secondly, using them to compare different processes, even fractional Gaussian noise with different Hurst indices, is too complicated a problem for an analytical solution [19]. The main difficulty is that the formula for the entropy of a Gaussian vector contains the determinant of the covariance matrix, the calculation of which there are no simple proposals for at the moment, except for cumbersome standard formulas, and at the same time, they do not make it possible to compare these determinants. Therefore, having at our disposal several classes of fractional processes that model a wide variety of processes, from physics to financial mathematics, we set out to compare in as simple a way as possible the terms of the information they carry, or, more simply, to compare their entropies. The comparisons of entropies presented are based on calculating the variances of the corresponding processes.

It is worth noting that certain theoretical properties, particularly the convexity of the Tsallis entropy, are challenging to analyze analytically. In such cases, we employ numerical investigations, which offer insights into theoretical properties.

III. RESULTS

We compare the entropies of the one-dimensional distributions of the following fractional processes: fractional Brownian motion, subfractional Brownian motion, Riemann–Liouville fractional Brownian motion, bifractional Brownian motion and three types of multifractional Brownian motion (moving-average, Volterra-type and harmonizable), as well as tempered fractional Brownian motions of the first and second kind.

We consider normalized versions of these processes to ensure their variances at $t=1$ to be equal 1. After this normalization, we observe that fractional Brownian motion, subfractional Brownian motion and Riemann–Liouville fractional Brownian motion share the same entropies. Similar formulas apply to bifractional Brownian motion; furthermore, its entropies can be compared to those of fractional Brownian motion depending on the values of t .

For multifractional Brownian motion, we have established that the moving-average and harmonizable versions of this process have the same entropies. These entropies can be compared with the corresponding entropies of Volterra-type multifractional Brownian motion, depending on the behavior of the Hurst function.

Lastly, for two versions of tempered fractional Brownian motions, we can numerically compare their entropies depending on the ratio between the multiplicative constants involved in their definitions.

IV. DISCUSSION

Our reason and goal of this comparison was to consider fractional processes from the point of view of quantity of information contained in their one-dimensional distributions, because these processes previously were mostly compared from the point of view of the behavior of their trajectories that is interesting in financial applications, but entropy properties are more interesting in

physical applications, for example, in the calculation of the fractal dimension of a solid sample. However, there is also an application to financial models. Namely, the Hurst index of fractional processes affects the behavior of their trajectories; its decrease leads to their irregularity and vice versa. But from the point of view of entropies, the situation showed a dependence on time: near zero, more precisely from zero to one, the variance, and therefore entropy, increases when the Hurst index decreases, but when time passes through unity, the situation changes to the opposite. This means that so-called rough volatility, which corresponds to the instability of the model, plays a crucial role only on short time intervals.

V. FUTURE RESEARCH

Our research opens up possibilities for future extensions in several directions. Potential avenues for further investigation include exploring various entropy measures for non-Gaussian processes, nonstationary processes and processes with nonstationary increments. Additionally, we can delve into the solutions of stochastic differential equations that describe the interactions of particle systems within random environments.

VI. CONCLUSION

We examined five distinct entropy measures applied to the Gaussian distribution: Shannon entropy, Rényi entropy, generalized Rényi entropy, Tsallis entropy and Sharma–Mittal entropy. We investigated their interrelationships and analyzed their properties in terms of their dependence on specific parameters. Our study extends to fractional Gaussian processes, encompassing fractional Brownian motion, subfractional Brownian motion, bifractional Brownian motion, multifractional Brownian motion and tempered fractional Brownian motion. We conducted a comparative analysis of the entropies associated with the one-dimensional distributions of these processes.

VIII. ACKNOWLEDGMENT

The second author is supported by The Swedish Foundation for Strategic Research, grant Nr. UKR22-0017 and by Japan Science and Technology Agency CREST, project reference number JPMJCR2115. The third author acknowledges that the present research is carried out within the frame and support of the ToppForsk project nr.274410 of the Research Council of Norway with title STORM: Stochastics for Time-Space Risk Models. The authors gratefully acknowledge the contributions of supporting organizations.

IX. DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," in *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, July 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x
- [2] B. Schneier, "Applied Cryptography: Protocols, Algorithms, and Source Code in C," Wiley, Hoboken, NJ, USA, 1996.
- [3] M. Nei and F. Tajima, "DNA polymorphism detectable by restriction endonucleases," *Genetics*, vol. 97, pp. 145-163, 1981.
- [4] W. Brock, J. Lakonishok, and B. LeBaron, "Simple technical trading rules and the stochastic properties of stock returns," *J. Financ.*, vol. 47, pp. 1731-1764, 1992.
- [5] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, pp. 130-141, 1963.
- [6] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge University Press, Cambridge, UK, 1994.
- [7] F. C. Morabito, D. Labate, and F. L. Foresta, "Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG," *Entropy*, vol. 14, pp. 1186-1202, 2012.

- [8] Y. Wu., P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R. M. Rangayyan, "Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters." *Biomed. Signal Process. Control*, vol. 31, pp. 265-271, 2017.
- [9] K. Li, W. Zhou, S. Yu, and B. Dai, "Effective DDoS Attacks Detection Using Generalized Entropy Metric." In *Proceedings of the Algorithms and Architectures for Parallel Processing; 9th International Conference, ICA3PP 2009, Taipei, Taiwan, June 8-11, 2009*; A. Hua and S. L. Chang, Eds., Springer, Berlin/Heidelberg, Germany, 2009, pp. 266-280.
- [10] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, pp. 81-106, 1986.
- [11] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Stat. Phys.*, vol. 52, pp. 479-487, 1988.
- [12] C. Tsallis, "The nonadditive entropy S_q and its applications in physics and elsewhere: Some remarks", *Entropy*, vol. 13, no. 10, pp. 1765–1804, Sep. 2011, doi: 10.3390/e13101765.
- [13] B. D. Sharma and I. J. Taneja, "Entropy of type (α, β) and other generalized measures in information theory," *Metrika*, vol. 22, pp. 205-215, 1975, doi: 10.1007/BF01899728
- [14] B. D. Sharma and D. P. Mittal, "New non-additive measures of relative information," *J. Combin. Inform. Syst. Sci.*, vol. 2, pp. 122-132, 1977.
- [15] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, University California Press, Berkeley, CA, USA; Los Angeles, CA, USA, June 20-July 30, 1960.
- [16] F. Nielsen and R. Nock, "A closed-form expression for the Sharma-Mittal entropy of exponential families," *J. Phys. A*, vol. 45, p. 032003, 2012.
- [17] F. Buryak and Y. Mishura, "Convexity and robustness of the Rényi entropy," *Mod. Stoch. Theory Appl.*, vol. 8, pp. 387-412, 2021.
- [18] R. L. Stratonovich, "Theory of Information and Its Value," Springer, Cham, Switzerland, 2020.
- [19] A. Malyarenko, Y. Mishura, K. Ralchenko, and Y.A.Rudyk, "Properties of Various Entropies of Gaussian Distribution and Comparison of Entropies of Fractional Processes". *Axioms*, vol. 12, no. 11, p. 1026, Oct. 2023, doi: 10.3390/axioms12111026

Застосування автоматизованих засобів для забезпечення прийняття рішень при віддаленому управлінні

Олег Посашков, Олександр Цимбал

*Кафедра комп'ютерно-інтегрованих технологій, автоматизації та робототехніки
Харківський національний університет радіоелектроніки.
Харків, Україна*

Анотація — У цьому тезисі розглядається використання автоматизованих засобів для прийняття рішень у віддаленому управлінні. Автори досліджують ефективність автоматизації, а також виявляють проблеми та перспективи впровадження цієї практики. Дослідження базується на аналізі літератури та використанні практичних прикладів.

Ключові слова — *Автоматизація, віддалене управління, ефективність, виклики, перспективи.*

I. ВСТУП

У світі віддалене управління стає дедалі важливішим елементом організаційної діяльності. Застосування автоматизованих засобів забезпечення прийняття рішень при віддаленому управлінні грає ключову роль забезпечення ефективності цього процесу. В данній роботі досліджується ефективність автоматизації у контексті віддаленого управління, а також аналізує перспективи розвитку цієї галузі.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Однією з ключових переваг автоматизації у віддаленому управлінні є здатність обробляти та аналізувати величезні обсяги даних. Сучасні системи штучного інтелекту та аналітичні інструменти дозволяють автоматично аналізувати інформацію та виявляти приховані закономірності [1]. Це істотно підвищує якість прийнятих рішень, особливо в умовах віддаленої роботи, де доступ до даних може бути обмежений. Застосування автоматизації у віддаленому управлінні також сприяє зниженню операційних витрат та ризиків. Автоматизовані системи можуть працювати цілодобово без перерв, що забезпечує безперервність бізнес-процесів. Більше того, зниження людського втручання може скоротити ймовірність людських помилок, що особливо важливо у критичних ситуаціях.

III. МЕТОДИ

Віддалене управління передбачає керування комп'ютерними системами та мережами з віддаленого місцезнаходження. Існує кілька методів, які дозволяють здійснювати віддалений доступ та управління:

- Віддалений робочий стіл (RDP) - дозволяє підключатися до віддалених комп'ютерів та працювати з ними в графічному інтерфейсі. Підтримується в ОС Windows.
- SSH - протокол зашифрованого віддаленого доступу, що використовується для текстового з'єднання та управління. Поширений в ОС Linux та UNIX.
- Віртуальні приватні мережі (VPN) - створюють зашифрований канал між пристроями та надають віддалений доступ до локальних мереж.

Для автоматизації процесів прийняття рішень можуть застосовуватися додаткові програмні рішення, такі як системи контролю версій (Git, SVN), CI/CD системи (Jenkins, TeamCity), системи моніторингу (Nagios, Zabbix).

Операційні системи Windows та Linux мають вбудовані механізми для організації віддаленого доступу та управління. Linux краще підходить для створення масштабованих інфраструктур завдяки відкритому вихідному коду, можливостям кастомізації та кращій продуктивності.

Для віддаленого керування роботизованими об'єктами, які розглядаються у дослідженні, обрано протокол SSH у поєднанні з бортовою системою комп'ютерного зору і датчиками відстані. Обрані засоби забезпечують також необхідні процеси прийняття рішень віддалено керованого мобільного робота.

Автоматизація забезпечує значне скорочення часу проведення оперативних дій [2], що призводить до підвищення продуктивності та мінімізації можливостей помилок персоналу. Додатково, автоматизація зменшує ризик людських помилок [3], особливо в умовах стресу та підвищеної складності. Автоматизовані системи діють індивідуально, гарантуючи високу точність та надійність. Розроблені рішення, базовані на автоматизації, дозволяють організаціям оперативно реагувати на зміни в швидкозмінному середовищі [4], а також адаптуватися до нових вимог та надавати аналітичні дані для стратегічного прийняття рішень.

Описані методи розглядаються з точки зору їхньої застосовності для створення ефективних систем віддаленого управління та забезпечення надійності та конфіденційності оброблюваної інформації.

IV. РЕЗУЛЬТАТИ

Аналіз методів віддаленого управління та можливостей їх автоматизації за допомогою програмних рішень показує, що існують ефективні способи організації безпечного віддаленого адміністрування та прийняття рішень.

Використання протоколів віддаленого доступу (RDP, SSH, VPN) в поєднанні з системами моніторингу, контролю версій, CI/CD дозволяє оптимізувати процеси управління IT інфраструктурою та розгортання програмного забезпечення.

Особливо ефективним є поєднання відкритих та кастомізованих рішень на базі ОС Linux для створення гнучких та масштабованих інфраструктур віддаленого управління.

V. ОБГОВОРЕННЯ

Однак, впровадження автоматизації у віддалене керування не без викликів та обмежень. Один з основних викликів – це питання безпеки даних та конфіденційності інформації [5]. Передача та зберігання чутливих даних можуть стати мішенню для кібератак та витоків інформації. Крім того, важливо враховувати, що не всі рішення та завдання можуть бути автоматизовані з рівною ефективністю. Наприклад, у деяких випадках, де потрібний високий рівень креативності чи морального розгляду, автоматизація може бути обмежена.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Перспективи використання автоматизованих засобів у віддаленому управлінні обіцяють бути дуже обнадійливими. Розвиток машинного навчання та штучного інтелекту сприятиме

створенню більш інтелектуальних та адаптивних систем для прийняття рішень [6]. Роботизований процесний автоматизації (RPA) буде відігравати все більш важливу роль в оптимізації бізнес-процесів. Майбутнє віддаленого управління, ймовірно, значною мірою визначатиметься тим, наскільки успішно організації зможуть використовувати ці технології для покращення своєї ефективності та конкурентоспроможності.

VII. ВИСНОВКИ

Використання автоматизованих засобів для прийняття рішень у віддаленому управлінні стає все більш важливим аспектом сучасні іноформаційні системи. Правильне використання автоматизації може підвищити ефективність організації, прискорити процеси та забезпечити більш гнучке та ефективне віддалене управління. Однак для досягнення максимальних результатів важливо враховувати виклики та слідкувати за розвитком технологій у цій галузі. Віддалене управління роботами і прийняття рішень реалізовано поєднанням SSH, комп'ютерного зору і датчиків відстані.

ЛІТЕРАТУРА

- [1] Горленко, С. В. "Автоматизація штучного інтелекту в управлінні бізнесом." Київський національний університет імені Тараса Шевченка, 2022.
- [2] Brynjolfsson, Erik, and Andrew McAfee. "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies." W. W. Norton & Company, 2016.
- [3] Davenport, Thomas H. "Process innovation: reengineering work through information technology." Harvard Business Press, 2013.
- [4] Smith, Aaron. "AI and machine learning in business: The state of play in 2019 and beyond." Pew Research Center, 2019.
- [5] Johnson, M., & Smith, H. J. (2019). Data breaches: Crisis and opportunity. *Business Horizons*, 62(1), 111-118.
- [6] Chui, M., Manyika, J., & Miremadi, M. (2016). Where machines could replace humans—and where they can't (yet). *McKinsey Quarterly*.

Intelligent decision support system based on recurrent neural networks and genetic algorithm for stock trading

Dmytro Uhryn, Andrii Bilyk

*Department of Information Technology
Yuriy Fedkovych National University
Chernivtsi, Ukraine*

Abstract—*The study develops a system that combines an artificial intelligence genetic algorithm (GA) with LSTM networks to optimise the prediction of OHLCV signals on the exchange. This system uses real-time data via web sockets, which leads to increased accuracy, efficiency and effectiveness of forecasting. This approach uniquely combines sequential machine learning techniques with a genetic algorithm, which marks a significant advance in financial forecasting modelling.*

Keywords—*Genetic Algorithms, LSTM Networks, Stock Prediction, OHLCV Signals, Real-time Data, Websockets.*

I. INTRODUCTION

With the advent of technology, the stock trading industry has undergone significant changes [1]. Gone are the days when traders relied solely on intuition and basic tools; in the data age, algorithmic trading based on artificial intelligence (AI) [2] is becoming the norm. Long Short-Term Memory (LSTM) networks [3] have become the preferred choice for processing sequential data, such as stock prices, due to their ability to remember patterns over long periods of time. However, training such deep learning models requires significant computational resources and expertise in tuning hyperparameters. Our research takes a deep dive into this problem by proposing an innovative solution. Using a genetic algorithm (GA), we have created an optimised system tailored to the nuances of the stock market. This system is further complemented by the use of real-time data via web sockets [4], which ensures timely and accurate forecasting in an ever-changing market environment.

II. RELATED WORK

In recent years, the domain of stock market prediction has witnessed a profound shift, primarily driven by the advancements in AI. With vast amounts of financial data being generated daily, there's a dire need for sophisticated methodologies capable of deciphering these intricate patterns. Among these methods, Long Short-Term Memory (LSTM) networks have emerged as one of the front runners [3]. Their inherent capability to deal with sequential data, remembering long-term dependencies, and avoiding the pitfalls of traditional Recurrent Neural Networks (RNNs) [5], such as the vanishing gradient problem, makes them aptly suited for the volatile world of stock market data. However, while LSTMs have shown promise, there's a burgeoning interest in enhancing their efficacy further [6-7].

III. METHODOLOGY

The designed system uses genetic algorithms to optimise LSTM networks for stock forecasting. This includes the following steps:

1. Data collection. It starts with collecting the necessary data for stock forecasting.

2. Data pre-processing. Pre-processing steps are performed to clean and refine the data, such as reducing noise in the data to ensure accuracy and normalising the data to bring all input characteristics to a standard scale, which will help in efficient model training.

3. Data partitioning. The data set is split: 82% is allocated for training purposes to train the model and 18% is reserved as test data to evaluate the model's performance on unseen data.

4. LSTM training:

4.1 Creating an initial population of LSTM models.

4.2. Training the LSTM model on the training data:

- The "fitness" of each model or the accuracy of stock price forecasting is evaluated.

- If the model is considered sub-optimal, it is transferred to genetic algorithm processes to improve it.

5. Processes of the genetic algorithm:

- Selection. Selection of models based on their suitability for participation in the crossover and mutation stages.

- Adaptive crossover. Combines the features of two parent models to create a new offspring model.

- Adaptive mutation. Making minor random changes to the model to find new potential solutions.

- After mutation, the new model is tested to see if it is optimal. If not, the process of selection, crossover and mutation is repeated until the optimal model is found.

6. Forecasting. The optimal LSTM model is used to predict stock closing prices.

7. Trend analysis. Based on the forecasted closing prices, the price trend is determined, indicating whether the stock is expected to rise, fall or remain stable.

8. Model evaluation. The efficiency of the model is evaluated using various metrics. An evaluation score is calculated to determine how closely the model's predictions match actual stock market trends.

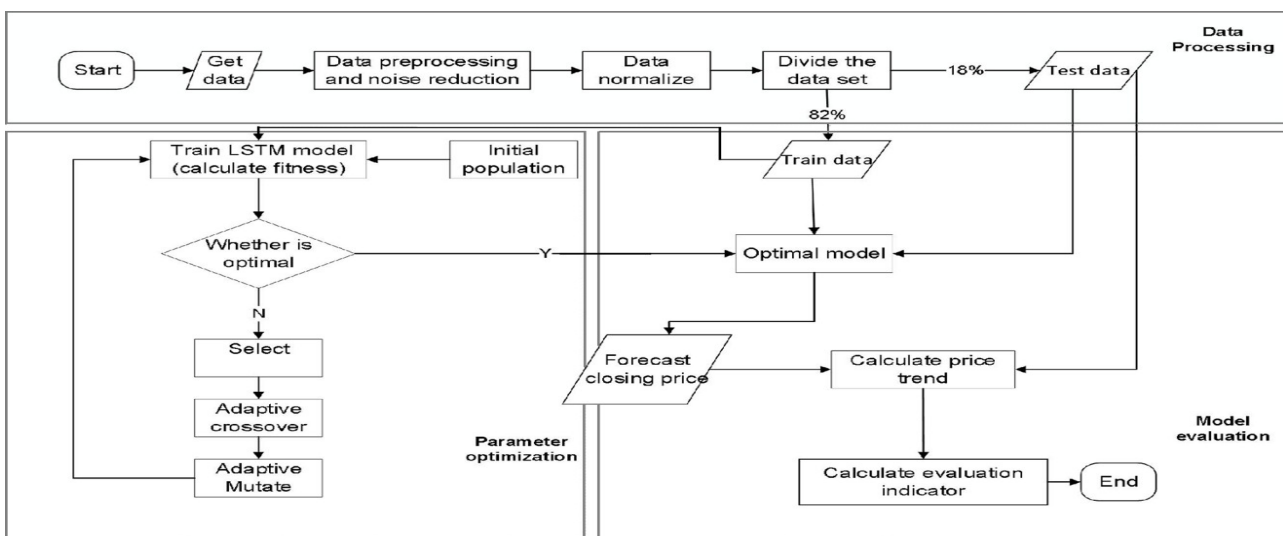


Figure 1. Scheme of architecture – interaction of GA and LSTM

Based on the developed approach, a system has been formulated that synergistically combines the evolutionary optimisation capabilities of a genetic algorithm with the advantages of sequential data processing of networks with long short-term memory LSTMs, which are adapted for accurate stock market forecasting.

IV. RESULTS

The proposed GA-optimised LSTM model will allow to obtain in automatic regime 55-58% accuracy rate in stock forecasting. The integration of real-time data via web sockets will further ensure timely and accurate forecasts in a dynamic stock market environment.

V. DISCUSSION

The study uses LSTM networks because of their ability to process time series data, which makes them effective for stock market forecasting. The main disadvantage of LSTM network is the need not only for knowledge of stock markets, but also deep knowledge of data science; in order to obtain results with 60% accuracy, you need to have in-depth knowledge. On the other hand, the genetic algorithm does not make it possible to complicate the model by adding layers of density, and in this case the user cannot go beyond 58% accuracy.

VI. FUTURE RESEARCH

Further research could delve into hybrid AI models and their application in various financial instruments.

VII. CONCLUSION

The combination of LSTM and GA-driven AI provides a robust system for stock prediction. With the integration of websockets, the system stays relevant with real-time data, showcasing its potential for real-world applications.

VIII. ACKNOWLEDGMENT

This study was carried out with the dedication and commitment of the research team. The authors extend their sincere appreciation to the Department of Data Science and Artificial Intelligence at Yuriy Fedkovych National University for their insightful guidance and profound expertise throughout the research process.

IX. DISCLOSURES

The authors declare that there are no financial affiliations or conflicts of interest related to the content presented in this paper. All research methodologies, data analyses, and conclusions were conducted independently and without external influence.

REFERENCES

- [1] S. Lee, J. Kim, H. Kang, D.-Y. Kang, and J. Park, "Genetic Algorithm Based Deep Learning Neural Network Structure and Hyperparameter Optimization," *Appl. Sci.*, vol. 11, no. 2, p. 744, Jan. 2021. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.3390/app11020744>.
- [2] G. P. Liu and V. Kadiramanathan, "Predictive control for non-linear systems using neural networks," *Int. J. Control*, vol. 71, no. 6, pp. 1119–1132, Jan. 1998. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.1080/002071798221515>.
- [3] S. Bakken, "Predictive models: important problems and innovative methods," *J. Amer. Med. Inform. Assoc.*, vol. 29, no. 1, pp. 1–2, Dec. 2021. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.1093/jamia/ocab274>.
- [4] "A Model for Stock Price Predictions Using Deep Learning Techniques," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 5, pp. 8266–8271, Oct. 2020. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.30534/ijatcse/2020/195952020>.
- [5] M. L. Chew and S. Puri, "Using Natural Language Processing Techniques for Stock Return Predictions," *SSRN Electron. J.*, 2017. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.2139/ssrn.2940564>.
- [6] S. Zhang, X. Ma, Z. Fang, H. Pan, G. Yang, and G. R. Arce, "Financial time series forecasting based on momentum-driven graph signal processing," *Appl. Intell.*, May 2023. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.1007/s10489-023-04563-y>.
- [7] L. Shen, H. Su, Z. Mao, X. Jing, and C. Jia, "Signal Property Information-Based Target Detection with Dual-Output Neural Network in Complex Environments," *Sensors*, vol. 23, no. 10, p. 4956, May 2023. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.3390/s23104956>.
- [8] W. C. and Mobile Computing, "Retracted: An Intelligent Evaluation Model of Enterprise Financial Data Quality Based on Artificial Neural Network," *Wireless Commun. Mobile Comput.*, vol. 2023, p. 1, July 2023. Accessed: Nov. 3, 2023. [Online]. Available: <https://doi.org/10.1155/2023/9874340>.

Використання гнучких комутаційних структур у складі апаратної частини мобільного робота

Ігор Невлюдов, Ірина Жарікова, Артем Бронніков

*Кафедра комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки
Харківський національний університет радіоелектроніки
Харків, Україна*

Анотація—Робота присвячена заміні жорстких компонентів апаратури на гнучкі з метою підвищення її якості, функціональних можливостей, надійності та зменшення габаритно-масових показників, що є особливо актуальним для мобільних роботів. Здійснено підбір компонентів і розроблено схему комутаційних зв'язків для мобільної платформи Keyestudio 4WD BT Robot car V2.0. Проведено моделювання комутаційної системи, зокрема виконано дослідження конструкції вузла комутаційних зв'язків для мобільного робота за допомогою середовища COMSOL Multiphysics.

Ключові слова—Автоматизація, між'єднання, поліїмідна структура, робот, комутаційна система, Arduino.

I. ВСТУП

Однією з основних тенденцій розвитку техніки є перехід від традиційної апаратури до електронних засобів на гнучких основах у рамках розвитку в усьому світі напряму гнучкої гібридної електроніки. В умовах мікромініатюризації електронних виробів застосування гнучких комутаційних систем (ГКС), до яких відносяться гнучкі електронні компоненти, елементи гнучкої електроніки, друковані плати й елементи між'єднань, забезпечує низку переваг під час створення як стаціонарних, так і рухомих конструкцій, зокрема комутаційної системи мобільних робототехнічних платформ. Тому дослідження у цій галузі сьогодні є вкрай актуальними [1-5].

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У компонентах гнучкої гібридної електроніки використовуються плівкові діелектричні матеріали, гнучкі електронні компоненти та рухомі конструкції пристроїв мікросистемної техніки, гнучкі та гнучко-жорсткі багатопарові друковані плати та шлейфи. У цьому застосуванні вони дозволяють вирішити питання більш щільного компонування вузлів апаратури, зниження габаритно-масових характеристик між'єднань.

Як діелектрична основа ГКС застосовуються найрізноманітніші полімерні матеріали: полієфіри, зокрема поліетилентерефталат, поліїмід, фторополімерні плівки, та інші. Для ГКС більшості груп електронних пристроїв використовується поліїмід – високотемпературний полімер. Завдяки використанню поліїмідних структур під час модернізації робототехнічних платформ, їх комутаційна система буде мати більшу хімічну стійкість, високу міцність, стійкість до корозійних процесів та гарні ізоляційні параметри основ друкованих плат і шлейфів [6].

Мобільні робототехнічні системи відрізняються наявністю рухомого шасі з автоматично керованими приводами. Вони можуть бути колісними, ходячими, колісно-ходячими і гусеничними. Часто мобільні роботи мають маніпулятори, завдяки яким можуть застосовуватися в польових умовах – для військової, медичної чи аграрної галузей, у космічних польотах, на промислових підприємствах, а також у будівництві та на транспорті. На виробничих підприємствах об'єктом робототехніки є створення автоматичних візків (робокарів), що рухаються в цеху за заданою програмою при керуванні від ЕОМ [7-12]. Програма може легко переналагоджуватися. У складі мобільних робототехнічних платформ також можуть бути бортові комп'ютери, в яких є ГКС.

Тож предметом дослідження є комутаційна система мобільної робототехнічної платформи на основі гнучких поліімідних структур.

III. МЕТОДИ

У ході дослідження виконувалася модернізація комутаційної системи для мобільної платформи Keystudio 4WD BT Robot car V2.0, на рис. 1, а представлена схема її комутаційних зв'язків [13].

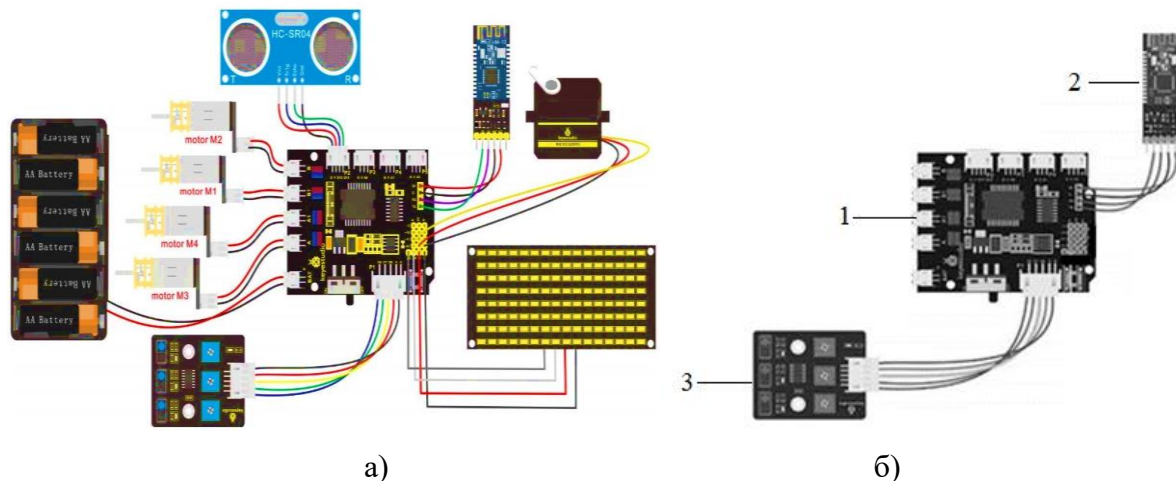


Рис. 1. Комутаційні зв'язки мобільної платформи Keystudio 4WD BT Robot car V2.0

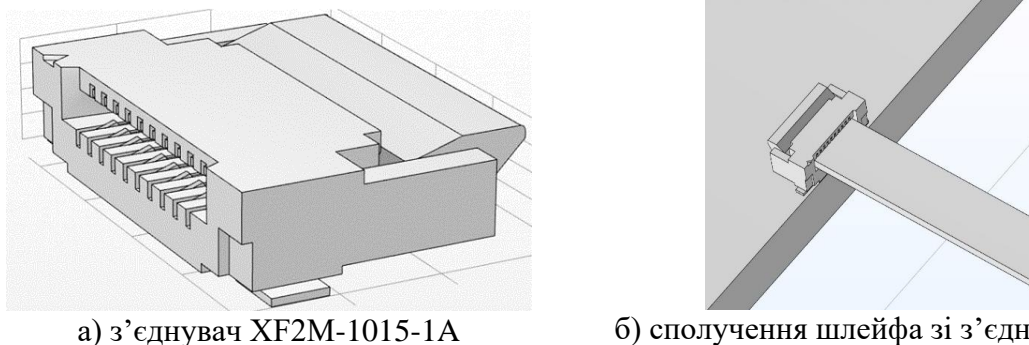
Для обґрунтування доцільності використання у складі мобільного робота обраних елементів між'єднань використано розрахунковий метод і метод комп'ютерного моделювання з використанням середовища COMSOL Multiphysics.

IV. РЕЗУЛЬТАТИ

Модернізацію комутаційної системи вказаного мобільного робота проводимо на основі заміни наведених на рис. 1, б проводів для з'єднання жорсткої друкованої плати 1 (модуль Keystudio Motor Driver Shield) з платами 2 (модуль Keystudio HM-10 Bluetooth-4.0) і 3 (модуль Keystudio Line Tracking Sensor) на гнучкі шлейфи з поліімиду марки ПІМ-А товщиною 100 мкм.

Досліджуваний гнучкий шлейф є системою мідних провідників, сформованих на гнучкій діелектричній плівці. Обраному матеріалу притаманні гарні фізичні характеристики, є збереження еластичності у широкому діапазоні температур. Основа маслостійка, не розчиняється в органічних розчинниках, помірно стійка до кислот і лугів, має високу радіаційну стійкість.

У середовищі COMSOL Multiphysics побудована модель вузла комутаційних зв'язків для мобільного робота Keystudio 4WD BT Robot car V2.0, модернізованого із застосуванням поліімідних ГКС (рис. 2).



а) з'єднувач XF2M-1015-1A
б) сполучення шлейфа зі з'єднувачем
Рис. 2. Імітаційна модель комутаційної системи мобільної робототехнічної платформи

У комутаційній системі використовуємо для з'єднання поліімідних шлейфів з жорсткими платами з'єднувачі XF2M з 10 виводами та кроком між ними 0,5 мм виробництва OMRON. Конектори цієї серії забезпечують надійну фіксацію шлейфа та контактний опір менше 100 мОм. Від окислення контакти з'єднувача захищає шар золота товщиною близько 0,12 мкм...0,15 мкм поверх нікелевого покриття 2 мкм.

V. ОБГОВОРЕННЯ

Для запропонованої системи міжз'єднань побудовано графік залежності опору стягування $R_{ст}$ від контактної натискання F_k (рис. 3).

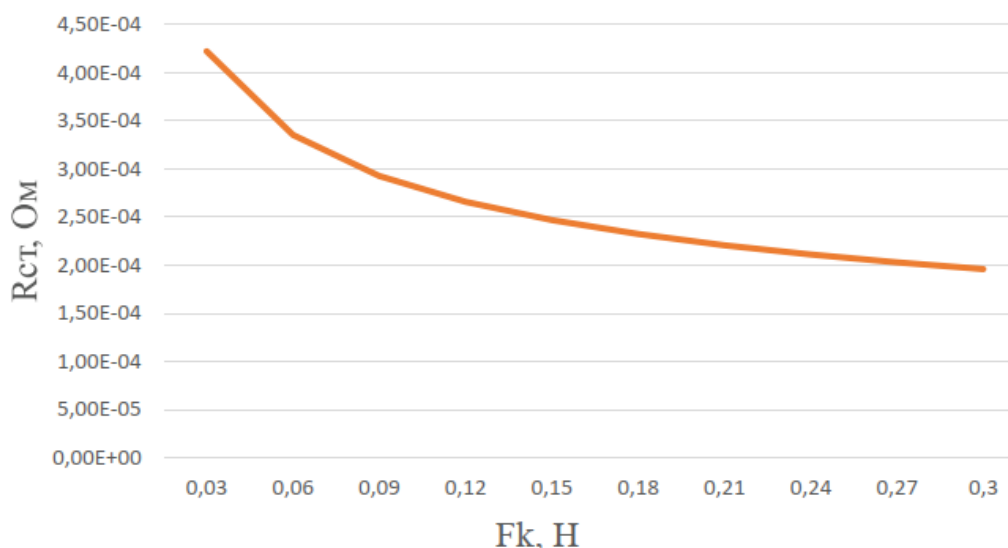


Рис. 3. Зміна опору стягування у контактах ГКС зі збільшенням сили натискання

Для визначення характеру залежності контактне натискання змінювалося у діапазоні 0,03...0,3 Н. Радіус площі контактної зони – 40 мкм. Отримані значення опору свідчать про припустиму якість комутаційного з'єднання у зоні контакту ГКС зі з'єднувачем.

Для моделювання механічних впливів на ГКС використовувався програмний комплекс COMSOL Multiphysics 4.2. Результат моделювання механічного впливу у 15 Н на викривлення ГКС показано на рис. 4. Отримані результати також підтверджують можливість використання ГКС у складі апаратної частини досліджуваного мобільного робота.

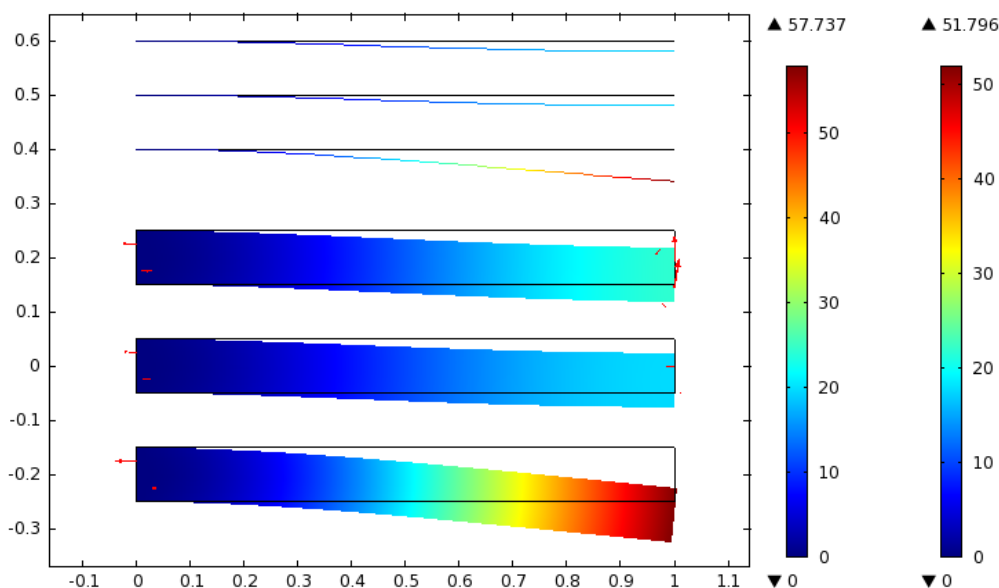


Рис. 4. Результат моделювання сили викривлення на гнучкий шлейф

Проведені на розривній машині дослідження показали, що розрив ГКС з фольгованого міддю полііміду відбувся при навантаженні 130 кгс та після подовження його на 14 мм. Також можна зробити висновок, що досліджений зразок витримує не критичні навантаження до 40 кгс без суттєвих деформацій структури ізоляційного та провідникового шарів. Також розвиток досліджень можливий у експериментальному відпрацюванні різних конструктивних рішень ГКС для різних видів мобільних роботів з метою обґрунтування доцільності їхнього використання у складі подібних конструкцій.

VI. ВИСНОВКИ

Таким чином, за результатами аналізу сучасних конструктивних рішень для мобільних роботів різного призначення було виконано модернізацію комутаційної системи для робототехнічної платформи Keyestudio 4WD BT Robot car V2.0 на основі використання поліімідних структур. Зокрема, здійснено підбір компонентів і розроблено схему комутаційних зв'язків. Проведено моделювання комутаційної системи, а саме виконано дослідження конструкції вузла комутаційних зв'язків для мобільного робота за допомогою середовища COMSOL Multiphysics.

Подальші дослідження будуть ґрунтуватися на виготовленні зразків ГКС і проведенні експериментального вивчення напружено-деформованого стану полііміду та провідникової системи з різних матеріалів – міді чи алюмінію – за різних зовнішніх механічних навантажень. Результати таких досліджень будуть спрямовані на підбір оптимальної топології ГКС і шляху її прокладання у конструкції мобільної роботизованої платформи.

ЛІТЕРАТУРА

- [1] Jonathan E, Jack C, Chang L. "Development of polyimide flexible tactile sensor skin", *J. Micromech. Microeng.* 2003; 13: 359-366. DOI: 10.1088/0960-1317/13/3/302.
- [2] Невлюдов І. Ш., Ботман І. В., Невлюдова В. В., Разумов-Фризюк Є. А. "Технологічне забезпечення якості гнучких комутаційних структур": монографія. – Кривий ріг: *KK HAV*, 2018. – 256 с. ISBN 980-969-578-288-9.
- [3] Palagin V., Razumov-Fryziuk I., Botsman I., Nevliudova V. "Development of multi-probe connecting devices on flexible polyimide base for MEMS components testing". in *14th International Conference on Perspective Technologies and Methods in MEMS Design, MEMSTECH*, 2018. DOI: 10.1109/MEMSTECH.2018.8365740.
- [4] Nevliudov I. Sh., Palagin V. A., Razumov-Frizjuk E. A., Zharikova I. V. "MEMS Intellect Multiprobes Contacting Devices for Electrical Checking-up of Multilayers Commutative Boards and BGA/CSP Electronic Components". *Proceedings of IEEE East-West Design & Test Symposium (EWDTS'2012)*. September 14-17, 2012. Kharkov, Ukraine. PP. 483-485.
- [5] Невлюдов І. Ш., Палагин В. А., Жарикова І. В. "Метод подключения электронных компонентов к автоматизированным измерительным комплексам", *Восточно-Европейский журнал передовых технологий*, vol. 1(9), pp.С. 4-7, 2013.
- [6] Nevliudov I., Zharikova I., Bronnikov A. "Improvement of the commutation system for a mobile robot platform using polyimide structures", *Eurasian scientific discussions. Proceedings of the 4th International scientific and practical conference (May 8-10, 2022)*, Barca Academy Publishing, Barcelona, Spain., pp. 157-163, 2022.
- [7] Tsybmal O., Bronnikov A., Gudkov Y. "Genetic Algorithm Solution for Transfer Robot Operation", *International Journal of Computing*, vol. 20(2), pp. 246-253, 2021. DOI: 10.47839/ijc.20.2.2172.
- [8] "A Mobile Robot Driven by Miniature Onboard", Accessed: Sep. 19, 2023 [Online]. Available: <https://www.medicaldevice-network.com/projects/mobile-robot/>.
- [9] Nevliudov I., Tsybmal O., Bronnikov A., Mordyk O. "Internet of things for robotic projects", *Сучасний стан наукових досліджень та технологій в промисловості*. № 3 (13). с. 58-64, 2020. DOI:10.30837/ITSSI.2020.13.058.
- [10] V. Bortnikova, V. Yevsieiev, V. Beskorovainyi, I. Nevliudov, I. Botsman and S. Maksymova, "Structural parameters influence on a soft robotic manipulator finger bend angle simulation", *15th Int. Conf. Exp. Des. Appl. CAD Syst. CADSM 2019*, pp. 35-38, 2019. DOI: 10.1109/CADSM.2019.8779300.
- [11] Igor Nevliudov, Oleksandr Tsybmal, Artem Bronnikov "Improvement of robotic systems based on visual control", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (18), pp. 65-74, 2021. DOI:10.30837/ITSSI.2021.18.065.
- [12] "Robotic Flexible Electronics with Self-Bendable Films", Accessed: Sep. 26, 2023 [Online]. Available: https://www.researchgate.net/publication/326561610_Robotic_Flexible_Electronics_with_Self-Bendable_Films. DOI: 10.1089/soro.2017.0141.
- [13] Ігнатенко Д. В. "Аналіз шляхів модернізації комутаційної системи мобільної робототехнічної платформи з використанням гнучких структур", *«Automation and Development of Electronic Devices» ADED-2021: збірник студентських наукових статей, ХНУРЕ*, Вип. 2, с. 58-62, 2021.

Метод аналізу супутникових знімків на основі глибинного навчання для застосування в обороні країни

Юрій Ушенко, Дмитро Угрин, Владислав Дашкевич

Кафедра комп'ютерних наук

Чернівецький національний університет імені Юрія Федьковича

Чернівці, Україна

Анотація—Проведено дослідження потенціалу використання технологій глибинного навчання для аналізу супутникових знімків у військовій сфері України. З огляду на постійні загрози безпеці і потребу у вдосконаленні оборонних можливостей, використання штучного інтелекту для обробки великої кількості даних і забезпечення точного розпізнавання об'єктів є критично важливим. В роботі представлені методи глибинного навчання, які можуть бути використані для ідентифікації військових об'єктів, переміщення та іншої важливої інформації, що виявляється на супутникових зображеннях.

Ключові слова — *Глибинне навчання, Супутниковий аналіз, Військові застосування, Штучний інтелект, Розпізнавання образів.*

I. Вступ

Сучасний світ пред'являє високі вимоги до національної безпеки та оборони, зокрема у контексті інформаційної війни та військових конфліктів. Україна, зіткнувшись з агресією, стоїть перед викликом забезпечення ефективної оборони та безпеки своєї території. Використання сучасних технологічних рішень, зокрема аналізу супутникових знімків за допомогою методів глибинного навчання, може істотно покращити можливості спостереження та реагування. В роботі досліджуються можливості застосування глибинного навчання для обробки та інтерпретації супутникових даних з метою ідентифікації потенційних загроз і підвищення точності військово-стратегічних рішень. Проведений аналіз сучасного стану галузі та запропоновано шляхи впровадження інформаційних технологій глибинного навчання для застосувань в оборонній індустрії України.

II. Аналіз останніх досліджень і публікацій

У сфері військового застосування супутникових знімків і глибинного навчання, останнє десятиліття характеризується значними досягненнями. Ранні роботи фокусувалися на базовому використанні супутникових даних для моніторингу географічних областей з метою розвідки і визначення військових об'єктів. Однак, із зростанням обсягів даних і появою потужніших алгоритмів глибинного навчання, дослідження перейшли до вдосконалення методів машинного зору і автоматизованого розпізнавання образів.

Однією з ключових робіт у цій області є стаття X. Zhou (2019), яка розглядає використання конволюційних нейронних мереж для ідентифікації військових об'єктів на супутникових зображеннях з високою точністю [1]. Їх підхід демонструє значне покращення у порівнянні з традиційними методами обробки зображень і служить основою для багатьох подальших досліджень.

На противагу цьому, праця E. Kalinicheva et al. (2020) зосереджена на використанні масштабованих нейронних мереж для аналізу часових рядів супутникових даних, дозволяючи виявляти та передбачати переміщення військових підрозділів з підвищеною точністю [2].

Нещодавно, зусилля дослідників були націлені на інтеграцію глибокого навчання із засобами штучного інтелекту для створення більш динамічних і адаптивних систем. Наприклад, робота V. N. Pandey et al. (2023) впроваджує інноваційну систему, яка використовує ансамблеві моделі глибокого навчання для обробки супутникових знімків, значно підвищуючи швидкість та точність виявлення змін у ландшафті [3].

III. Методи

У цьому дослідженні увагу зосереджено на використанні конволюційних нейронних мереж (CNN) для ідентифікації військових об'єктів на супутникових знімках. Цей підхід вибрано з огляду на його здатність автоматично виділяти важливі риси із зображень, що є критично важливим для точного розпізнавання об'єктів у різних умовах освітлення та приховування.

Для вирішення поставленої задачі обрано кількісний дослідницький підхід. Це дозволило зібрати об'єктивні дані та аналізувати їх за допомогою статистичних методів.

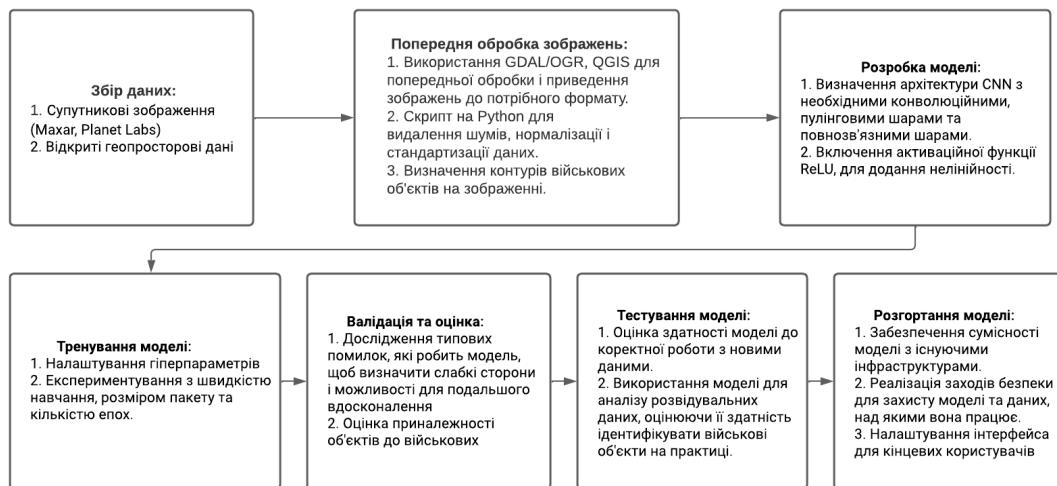


Рисунок 1. Послідовність етапів для побудови CNN моделі аналізу супутникових знімків на основі глибокого навчання

Для цього дослідження було спроектовано архітектуру CNN, яка включає в себе спеціалізовані модулі, базовані на принципах глибокого навчання. Ця архітектура була створена з метою підвищення точності класифікації об'єктів, особливо в ситуаціях, коли зображення характеризуються високим рівнем невизначеності, наприклад, через погану видимість, зміну освітлення або маскуванню об'єктів. Ця архітектура CNN була розроблена з урахуванням декількох ключових характеристик:

1. Багатошаровість: Складається з декількох шарів, включаючи конволюційні шари, пулінг-шари, і повнозв'язні шари, що дозволяє ефективно виявляти і абстрагувати особливості зображення на різних рівнях.

2. Глибинне навчання: Використовує багатошарову CNN, яка автоматизує виявлення характеристик об'єктів на зображеннях. Він впроваджує послідовність фільтрів, що самостійно адаптуються під час тренування, забезпечуючи високу точність розпізнавання.

3. Використання ReLU (rectified linear unit) як функцію активації для забезпечення нелінійності в моделі, що дозволяє мережі вчитися більш складних представлень даних.

4. Застосування метода регуляризації Dropout для уникнення перенавчання моделі і підвищення її здатності до узагальнення.

Для тренування та валідації CNN був використаний самостійно розроблений датасет супутникових зображень. Цей датасет містить близько 1000 зображень, кожне з яких було анотоване експертами. Кожне зображення було стандартизоване до розміру 256x256 пікселів, щоб забезпечити однакові умови для обробки всіма шарами CNN.

Датасет включає 5 основних класів військових об'єктів: літаки, гелікоптери, військові транспортні засоби, будівлі та стартові комплекси. Було зроблено зусилля для збалансування кількості зразків у кожному класі, щоб уникнути упередженості в навчанні. Зображення для датасету були зібрані з різних супутникових джерел, забезпечуючи різноманітність умов зйомки. Датасет було розділено на тренувальний (80%), валідаційний (10%) та тестовий (10%) набори.

Валідація моделі проводилася з використанням окремого набору зображень, що дозволило перевірити здатність моделі до узагальнення на нових даних.

Встановлено наступні критерії для оцінки ефективності моделі:

1. Точність: відсоток правильно ідентифікованих об'єктів.
2. Чутливість: спроможність моделі правильно ідентифікувати позитивні випадки.
3. Специфічність: здатність моделі правильно відкидати негативні випадки.
4. Обґрунтування методів.

Обрані методи обґрунтовані потребою в автоматизації процесу ідентифікації об'єктів та необхідністю впоратися з великою кількістю даних, які постійно надходять з супутникових джерел. Конволюційні нейронні мережі є передовою технологією в області комп'ютерного зору, що підтверджено численними дослідженнями, які демонструють їх високу ефективність у подібних завданнях.

Для забезпечення валідності та надійності результатів застосовано крос-валідацію та додаткову перевірку адекватності моделі з використанням незалежних даних. Також використано техніки збільшення даних для запобігання перенавчанню моделі та забезпечення її стійкості до різноманітності умов зображень.

IV. РЕЗУЛЬТАТИ

За допомогою раніше описаного методу, було досягнуто точності ідентифікації військових об'єктів на супутникових зображеннях на рівні 80%. Це становить покращення порівняно з традиційними методами, де середня точність складала близько 70-75%.

Таке значне підвищення точності демонструє ефективність застосування CNN у даній області та високу адаптивність до різних умов освітлення та камуфляжу, підвищуючи чутливість та специфічність системи. Валідація моделі на незалежному

наборі даних підтвердила її здатність ефективно узагальнювати навчання, відкриваючи шлях для подальшого розвитку інструментів супутникової розвідки.

Проте, результати також виявили деякі обмеження. Зокрема, модель показала точність (ассигасу) приблизно 80%, що, хоча й є покращенням порівняно з традиційними методами, все ж залишає простір для подальших вдосконалень. Ключовими показниками якості, такими як precision та recall, були відповідно 78% та 75%. Ці результати вказують на те, що модель ефективно ідентифікує більшість військових об'єктів, але все ще має труднощі з деякими видами камуфляжу або обмеженнями освітлення.

Таким чином, хоча модель і демонструє значний прогрес у сфері автоматизованого розпізнавання об'єктів на супутникових знімках, все ще існує потреба в подальших дослідженнях і вдосконаленнях, зокрема у виправленні помилок, пов'язаних з обмеженими умовами візуалізації та складністю об'єктів.

V. ОБГОВОРЕННЯ

Отримані результати демонструють важливість інтеграції новітніх технологій глибинного навчання у сфері військового супутникового спостереження. Значне підвищення точності ідентифікації вказує на потенціал застосування цих технологій для поліпшення розвідувальної інформації та оперативного реагування. Однак, існують обмеження, пов'язані з обробкою великих обсягів даних та необхідністю подальшої адаптації алгоритмів для конкретних сценаріїв використання. Важливо зазначити, що це дослідження є кроком до розробки більш складних і автономних систем спостереження.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У майбутньому дослідження можуть бути спрямовані на оптимізацію алгоритмів для прискорення обробки даних у режимі реального часу, а також розвитку методів машинного навчання для кращого розуміння контексту зображень. Це дозволить не лише виявляти, а й прогнозувати потенційні загрози.

VII. ВИСНОВКИ

Дослідження підтвердило гіпотезу про ефективність використання конволюційних нейронних мереж у військовому супутниковому розвідуванні. В ході дослідження було продемонстровано, що цей підхід може істотно підвищити точність ідентифікації об'єктів. Ключові аргументи, наведені в дослідженні, включають значне покращення точності ідентифікації, здатність до швидкої адаптації до нових умов та потенціал для подальшої оптимізації. Отже, використання глибинного навчання в супутникових системах військового застосування відкриває нові можливості для розвідки та безпеки.

ЛІТЕРАТУРА

- [1] X. Zhou et al., "Applying convolutional neural network for military object detection on embedded platform" in *Communications in computer and information science*, 3rd ed., vol. 1897, Springer, 2019, pp. 265–312.
- [2] E. Kalinicheva et al., "Unsupervised satellite image time series analysis using Deep Learning Techniques" dissertation, SORBONNE UNIVERSITY, 2020, p. 202.
- [3] B. N. Pandey et al., "Deep Learning based ensemble model for Satellite Image Classification", in *Computer Science*, vol. 1115, 2023, pp 24-28.

Analysis of a fixed-width binning method

Olga Solovei

*Department of Information Technology and Applied Mathematics
Kyiv National University of Construction and Architecture
Kyiv, Ukraine*

Abstract—The subject matter of the article is fixed-width binning method with constant number of bins. In this paper the focus is put on impact's analysis of "fixed-width binning" method with ten intervals on the ROC convex hull when the method is used to calibrate the predicted binary probabilities. The goal of work is study and prove that the method may lead to inconsistencies between ROC analysis's results received on ROC curve. To meet the goal in the research were used the methods: analysis of ROC construction algorithm and "fixed-width binning" method, experimental tests. It is concluded that fixed-width binning method with ten bins may calibrate probabilities so that they are "too-smoothed" which causes ROC convex hull is changed. The change in ROC convex hull leads to inconsistencies in ROC analysis results. In the paper is recommended to calculate number of bin for "fixed-width binning" method based on simple statistics of predicted uncalibrated binary probabilities.

Keywords—*ROC curve, fixed-width binning, calibrated probabilities, true positive rate, false positive rate, threshold.*

I. INTRODUCTION

Calibrated probabilities are needed for ROC analysis when it is conducted with one of the goals: 1) to select regions where one machine learning classifier outperforms another; 2) to identify the threshold that minimizes an error rate or for the given error rate to get the corresponding threshold and use it to predict a target class [1].

In binary classification task the probability is considered as well calibrated when it's value for positive class is closed to 1 and for negative - is closed to 0. Such estimates for probability are returned by Logistic regression classifier and Support Vector Machine Classifier (SVC), when SVC object is created with a directive to calibrate probabilities. The probabilities which are estimated by machine learning classifiers such as Random Forest Classifier (RFC) or Naive Bayes Classifier (GaussianNB) are required to be calibrated by one of the methods: "Platt calibration", "Isotonic regression" or "fixed-width binning".

"Platt calibration" and "Isotonic regression" methods are powerful calibration algorithms, however, when may cause the following issues [2,3,4]:

- "Platt calibration" method: 1) it is the most efficient when distortion of predicted scores is sigmoid-shaped; 2) if the data distribution changes over time then the coefficients of "Platt calibration" may become incorrect and as a result the probabilities are not calibrated.
- "Isotonic regression" method may not be efficient for imbalanced dataset and can be computationally intensive when dataset is large.

In research [5] was proved the efficiency of "fixed-width binning" method for classification algorithms Random Forest Classifier (RFC) and Naive Bayes Classifier (GaussianNB) and recommended to partition an interval [0..1] in ten bins.

However, a histogram to model a statistical properties must have a number of bins depending on the actual data [6,7], therefore a division the interval [0..1] in ten bins for "fixed-width binning" method to scale binary probabilities may not help to improve the reliability of predicted probabilities.

The current research objectives are to empirically study and prove that a "fixed-width binning" method to scale binary probabilities with ten intervals may lead to inconsistencies between ROC analysis's results received on ROC curve constructed with uncalibrated predicted probabilities and ROC curve constructed with calibrated probabilities and to recommend a direction on how the inconsistencies potentially to be avoided.

II. RELATED WORK

To calibrate probabilities which are predicted by classifiers: Random Forest Classifier (RFC) and Naive Bayes Classifier (GaussianNB) in study [8] is proposed a scaling binning method. The method improves “fixed-width binning” by calculating of the bin’s number depending on unscaled predicted scores. The algorithm divides data into two subsets: probabilities which are predicted for the first subset is calibrated by “Platt calibration” method and the probabilities which are predicted for the second subset are used to choose the bins so that an equal number of points are landed in each bin. The scaling binning method addresses two issues: reduces a calibration error and calculates bin width which is adopted to already calibrated scores. However, as the algorithm uses “Platt calibration” its results may be impacted by already mentioned problems associated with “Platt calibration”.

The related research [9] proves that the histogram binning method calibrates predicted probability well, providing the number of bins is chosen appropriately, however the research doesn’t focus on impact’s analysis of “fixed-width binning” with ten intervals on the ROC analysis results.

So, it is an actual task to study the impact of “fixed-width binning” with ten intervals on the ROC analysis results and make the recommendations on how the impact can be reduced.

III. METHODOLOGY

To achieve the study’s goals, we will use the following methodology:

- Construct ROC convex hulls for unscaled and scaled predicted probabilities.
- Compare classifier’s performance based on ROC convex hulls and record the difference.
- If the differences are found, then search for the reasons by analyzing the algorithm for ROC convex hull’s construction.

The methodology to be executed through the steps: 1). Generate input dataset to include 10 positive and 10 negative instances with predicted class and corresponding uncalibrated probability. 2). Calibrate probabilities with “fixed-width binning”. 3). Construct ROC convex hulls with calibrated and not calibrated probabilities. 4). Fix the desired number of misclassified negatives and record the required threshold on each ROC convex hull. 5). Compare the result and make the recommendations.

IV. MATERIALS

ROC curve depicts relation between true positive rate, defined according to equation 1, and false positive rate (equation 2) depending on chosen value of threshold, where threshold defines the minimum value of probabilities for ROC curve to indicate a positive instance.

$$TPR = \frac{\text{positive instance correctly classified}}{\text{total positives instances}} \quad (1)$$

$$FPR = \frac{\text{negative instance incorrectly classified}}{\text{total negative instances}} \quad (2)$$

Plot tpr on the Y axis and fpr are on the X axis enables to perform the analysis in order to select that value for probabilities’ threshold which will guarantee a classification algorithm’s performance. For example, in table 1 for 10 positive and 10 negative instances is specified a predicted class with corresponding uncalibrated probability. Corresponding ROC curve on picture (a) from figure 1 at threshold equal to 0.5 shows tpr is 0.7 (i.e., 7 positive instances were predicted correctly) and fpr is 0.3 (i.e. 3 negative instances were classified incorrectly). When a classifier is required to have less than 3 misclassified negatives then its threshold must be set to 0.51; when it is required to have less than 2 mistakes in positive class then its threshold must be 0.28. Thus, varying a threshold on ROC curve graph a desired performance of classifier is achieved.

Table1. Predicted an instance’s class with corresponding predicted unscaled probability.

score	.9	.8	.75	.6	.58	.54	.53	.52	.51	.5	.4	.3	.28	.27	.26	.25	.24	.23	.15	.1
Predicted class	1	1	0	1	1	1	0	1	1	0	1	0	1	0	0	0	1	0	1	0

“Fixed-width binning” method with constant number of bins equal to 10 divides interval [0..1] and for each predicted uncalibrated probability identifies its index of bin; then the probabilities which

are in the same bin got calibrated by average from uncalibrated probabilities in the same bin. As a result, calibrated probabilities in the same bin got the same value, i.e. there is no discrimination in values of scaled probabilities in the same bin. When ROC curve is constructed for the sorted calibrated probabilities, denoted as “calibrated_scores”, then only the 1st from the range of duplicated probabilities is took (line 6 in algorithm 1)) and as a result a value of true positive rate is decreased as line 9 is not hit.

Algorithm 1. ROC curve construction

Input: calibrated_scores, y_pred, n, p

Output: curve

1. y_pred_sorted=sort y_pred by calibrated scores in descending order
2. FP=TP=i=0
3. calibrated_scores_prev =inf
4. curve=[]
5. while i<|y_pred_sorted| do
6. if calibrated_scores[i]<> calibrated_scores_prev then
7. curve.add (FP/N, TP/P, calibrated_scores[i])
8. calibrated_scores_prev = calibrated_scores[i]
8. endif
9. if calibrated_scores[i]=1 then tp++
10. else fp++
11. endif
12. i++
13. endwhile

The reduced value of tpr as a result of calibrated score’s are “too-smoothed” leads to changes in ROC convex hull and as a result to of that change to inconsistencies in required threshold’s value for the same requirements to classifier’s performance when probabilities are calibrated and not calibrated.

V.RESULTS

The scores from Table 1 after calibrated by “fixed-width binning” method with ten intervals are included in Table 2. Corresponding ROC curve on picture (b) of figure 1 shows that a threshold equal to 0.75 must be set in order to have less than 3 misclassified negatives and a threshold equal to 0.28 is required to have less than 2 mistakes in positive class.

Table2. Predicted an instance’s class with corresponding predicted scaled probability.

score	.9	.8	.75	.54	.54	.54	.54	.54	.54	.54	.4	.26	.26	.26	.26	.26	.26	.125	.125	
Predicted class	1	1	0	1	1	1	0	1	1	0	1	0	1	0	0	0	1	0	1	0

VI.DISCUSSION

From the results is visible that ROC analysis results for the same requirements to classifier performance are different for graphs (a) and (b) from figure 1. To achieve the same performance to predict negative instances the threshold equals to 0.51 is needed, however it’s values is changed to 0.75 after probabilities are calibrated. Similar, to achieve the same performance to predict positive instances the threshold equals to 0.26 is changed to 0.28.

To avoid inconsistencies in ROC curve analysis results after binary predicted probabilities are calibrated by “fixed-width binning” it is recommended to decide the number of intervals considering statistical characteristics of predicted uncalibrated probabilities.

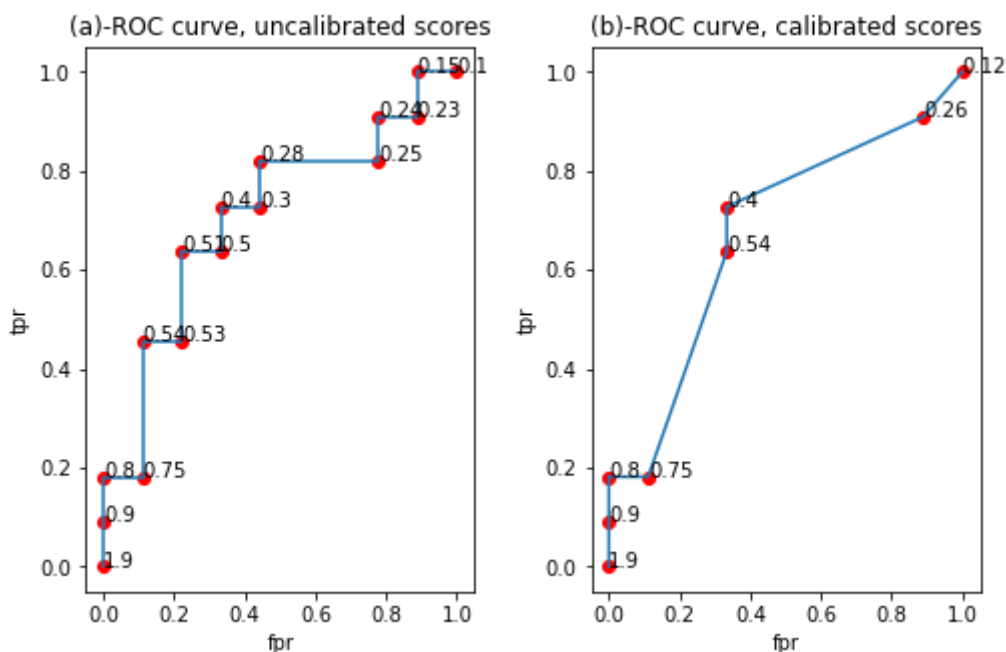


Figure 1. ROC convex hull for unscaled and scaled predicted probabilities.

VII.FUTURE RESEARCH

Future studies will be to propose an approach to calculate bin number for “fixed-width binning” based on simple statistics of predicted uncalibrated binary probabilities.

VIII.CONCLUSION

In this paper is specified that the scaled predicted probabilities with “fixed-width binning” and ten intervals may become “too-smoothed”, i.e. got assigned equal values. The logic of ROC curve construction algorithm removes duplicates in sorted list of calibrated scores which changes ROC convex hull and the change in convex hull causes inconsistencies between ROC analysis’s results received on ROC curve constructed with uncalibrated predicted scores and ROC curve constructed with calibrated scores. To avoid the mentioned impact on ROC convex hull it is recommended to calculate number of bins for “fixed-width binning” based on simple statistics of predicted uncalibrated binary probabilities.

REFERENCES

- [1] Fawcett, T., 2006. An introduction to ROC analysis. *Pattern recognition letters*, 27(8), pp.861-874.
- [2] Niculescu-Mizil, A. and Caruana, R., 2005, August. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning* (pp. 625-632).
- [3] Stout, Q.F., 2013. Isotonic regression via partitioning. *Algorithmica*, 66, pp.93-112.
- [4] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*. 1999 Mar 26;10(3):61-74.
- [5] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. *InIcml 2001 Jun 28 (Vol. 1, pp. 609-616)*.
- [6] Heinrich, C., 2021. On the number of bins in a rank histogram. *Quarterly Journal of the Royal Meteorological Society*, 147(734), pp.544-556.
- [7] D. Freedman D, P. Diaconis, On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. 1981 Dec;57(4):453-76.
- [8] Naeini, M.P., Cooper, G. and Hauskrecht, M., 2015, February. Obtaining well calibrated probabilities using bayesian binning. In *Proceedings of the AAAI conference on artificial intelligence (Vol. 29, No. 1)*.
- [9] Gupta, C. and Ramdas, A., 2021, July. Distribution-free calibration guarantees for histogram binning without sample splitting. In *International Conference on Machine Learning* (pp. 3942-3952). PMLR.

Структура матриць системи лінійних алгебраїчних рівнянь для задачі моделювання масопереносу в пористому середовищі на графі

Валерій Колесников

*Кафедра прикладної математики
Факультет комп'ютерних наук та кібернетики
Київський національний університет імені Тараса Шевченка
Київ, Україна*

Анотація—Дана робота містить аналіз структури матриць систем лінійних алгебраїчних рівнянь, що використовуються в чисельних методах при знаходженні наближеного розв'язку рівняння Річардса-Клюта на графі, яке описує процес масопереносу у системі капілярних труб. Також робота містить опис ефективних методів для розв'язку даної системи лінійних алгебраїчних рівнянь.

Ключові слова—*Математичне моделювання, Рівняння Річардса-Клюта, Графи, Система лінійних алгебраїчних рівнянь, Чисельні методи.*

I. ВСТУП

Процес масопереносу в пористих середовищах з межею насичення описується зазвичай за допомогою рівняння Річардса-Клюта. Дане рівняння є нелінійним еліптико-параболічним рівнянням у часткових похідних, через що відомо дуже мало аналітичних розв'язків цього рівняння, більшість з яких отримані для ідеалізованих параметрів середовища. Це є також причиною невеликої кількості апріорних оцінок для розв'язків рівняння, значна кількість з яких сформульована для слабких розв'язків. Основним інструментом для отримання наближених розв'язків рівняння Річардса-Клюта та моделювання масопереносу в пористому середовищі є чисельні методи. Проте за рахунок нелінійності процес розв'язання рівняння за допомогою чисельних методів займає дуже багато часу, тому актуальною проблемою є підвищення ефективності цього процесу. Зазвичай чисельні методи зводять задачу пошуку наближеного розв'язку рівняння на наступному кроці по часу до задачі знаходження розв'язку системи лінійних алгебраїчних рівнянь (СЛАР). Дана робота містить аналіз структури матриці СЛАР для рівняння Річардса-Клюта, що моделює процес масопереносу в системі капілярних труб, що з'єднані між собою. Ця система труб моделюється за допомогою графа, що вкладений у тривимірний простір. Також у роботі наведені ефективні методи розв'язання СЛАР, виходячи з результатів аналізу її структури.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Зазвичай у якості чисельних методів для розв'язання рівняння Річардса-Клюта використовують методи скінченних елементів, скінченних об'ємів та скінченних різниць зі зворотним сталим кроком по часу. Проблемі підвищення ефективності процесу моделювання масопереносу в пористому середовищі присвячена низка робіт. Серед них можна виділити роботу [1], в якій запропоновано адаптивний крок по часу, а також роботу [2], де описана процедура наближення розв'язку рівняння Річардса-Клюта, розроблена спеціально для неоднорідних середовищ. Окремо можна розглянути також чисельні методи, що

використовують спрощення рівняння Річардса-Клюта. Так, у роботі [3] наведені чисельні методи для лінеаризації рівняння Річардса-Клюта, а у роботі [4] пропонується стійка схема для апроксимації розв'язку рівняння Річардса-Клюта з модифікованим представленням потенціалу тиску. Огляд сучасних чисельних методів для розв'язання рівняння Річардса-Клюта можна знайти у роботах [5-6].

III. МАТРИЦЯ СЛАР

Для опису процесу масопереносу у системі капілярних труб можна використовувати тривимірне рівняння Річардса-Клюта, проте в такому випадку кількість вузлів в просторі, в яких знаходиться значення наближеного розв'язку за допомогою чисельного методу, буде дуже великим, і тим самим процес знаходження наближеного розв'язку буде малоєфективним. Але, якщо діаметри труб нехтовно малі в порівнянні з їх довжинами, то, припускаючи, що значення розв'язку в певній точці труби не залежить від положення точки на поперечному перерізі труби, то можна описати таку систему труб за допомогою вкладеного у тривимірний простір графа. Тоді для опису масопереносу в окремій трубці буде використовуватися одновимірне рівняння Річардса-Клюта, а у вершинах графу повинна виконуватися рівність, що описує баланс потоку маси через дану вершину. Вона буде мати наступний вигляд.

$$\sum_{e \in Inc(v)} q_e = 0, \quad (1)$$

де q_e — потік маси до вершини v з ребра e , що інцидентне даній вершині графа.

Тоді, якщо використовувати чисельні методи для одновимірних рівнянь Річардса-Клюта на ребрах, то при дискретизації рівності (1), як і при дискретизації крайової умови другого роду у випадку звичайного одновимірного рівняння, будуть використовуватися лише два крайніх вузли просторової сітки на кожному з ребер. Оскільки у внутрішніх вузлах ребер рівняння апроксимується повністю аналогічно до одновимірного випадку, то матриця СЛАР, що буде утворюватися на кожному кроці по часу, буде мати наступну структуру.

$$A = \begin{pmatrix} a_i & b_i & & & & & & & & d_i \\ c_i & a_i & b_i & & & & & & & \\ & c_i & a_i & b_i & & & & & & \\ & & c_i & a_i & b_i & & & & & d_i \\ & & & c_i & a_i & b_i & & & & \\ & & & & c_i & a_i & b_i & & & \\ & & & & & c_i & a_i & b_i & & \\ d_i & & & & & & c_i & a_i & b_i & \\ & & & & & & & c_i & a_i & b_i & d_i \\ & & & & & & & & d_i & c_i & a_i & b_i \\ & & & & & & & & & & c_i & a_i & b_i \\ & & & & & & & & & & & d_i & c_i & a_i \end{pmatrix}. \quad (2)$$

Тридіагональні блоки матриці відповідають апроксимації рівняння Річардса-Клюта на кожному з ребер, в той час як інші рядки матриці, в яких більше трьох ненульових елементів, відповідають рівнянню балансу потоків маси (1) і кожен з цих рядків відповідає деякій вершині графа. Таким чином, матриця (2) СЛАР, що утворюється при використанні

чисельних методів для розв'язання рівняння Річардса-Клюта на графі, має наступні властивості:

- 1) A — розріджена,
- 2) A — симетрична,
- 3) A має діагональну перевагу,
- 4) кількість рядків у матриці A з більш ніж трьома ненульовими елементами не залежить від параметрів дискретизації рівняння, тобто не залежить від розміру самої матриці.

У випадку тривимірної апроксимації матриця СЛАР має властивості 1)-3), за рахунок чого для розв'язання СЛАР можна використовувати модифіковані для розріджених матриць методи Зейделя та спряжених градієнтів. Проте, оскільки для рівняння Річардса-Клюта на графі виконується також умова 4), то, на додачу до цих методів, можна запропонувати використовувати метод Гауса, модифікований для розріджених матриць. Тоді, після прямого проходу методу Гауса, що займає час $O(n)$, де n — кількість рядків у матриці A , матриця A набуде наступного вигляду.

$$A' = \begin{pmatrix} a'_i & b'_i & & & & & & & d'_i \\ & a'_i & b'_i & & & & & & d'_i \\ & & a'_i & b'_i & & & & & d'_i \\ & & & a'_i & b'_i & & & & d'_i & d'_i \\ & & & & a'_i & b'_i & & & d'_i & d'_i \\ & & & & & a'_i & b'_i & & d'_i & d'_i \\ & & & & & & a'_i & b'_i & d'_i & \\ & & & & & & & a'_i & b'_i & d'_i \\ & & & & & & & & a'_i & b'_i & d'_i \\ & & & & & & & & & a'_i & b'_i \\ & & & & & & & & & & a'_i \end{pmatrix}. \quad (3)$$

Зворотний хід метода Гауса також займатиме час $O(n)$, тому і весь час, витрачений на розв'язання СЛАР, можна оцінити як $O(n)$, причому точніше його можна оцінити як $|V| \cdot n$, де V — множина вершин графу. Варто зазначити, що даний метод дає точний розв'язок СЛАР.

IV. РЕЗУЛЬТАТИ

Для рівняння Річардса-Клюта на графі ми отримали структуру матриці СЛАР, що виникає на кожному кроці по часу при використанні чисельних методів для знаходження наближеного розв'язку рівняння. З властивостей цієї структури були визначені методи розв'язання СЛАР, що мають найбільшу ефективність в даному випадку, серед яких: метод Зейделя, метод спряжених градієнтів, метод Гауса.

V. ОБГОВОРЕННЯ

Дані результати стосовно структури матриці СЛАР, що утворюється при використанні чисельних методів для розв'язання рівняння Річардса-Клюта на графі, вказують на те, що для

ефективного отримання розв'язку СЛАР можна не обмежуватися стандартними методами Зейделя та спряжених градієнтів, які активно використовуються у дво- та тривимірних випадках, а розробляти нові методи та модифікації вже існуючих методів, які, можливо, даватимуть точний розв'язок СЛАР та працюватимуть за лінійний час. Одним з таких методів, як і було вказано в роботі, є модифікований для розріджених матриць метод Гауса. Основним аргументом для доведення ефективності даного методу була властивість 4) матриці СЛАР, що дозволила отримати лінійну залежність від розмірності матриці для часу, що витрачається на виконання модифікованого методу Гауса.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У якості подальших напрямків досліджень можна розглянути питання модифікації методу прогонки для матриць зі структурою, подібною до структури (2), де ненульові елементи поза трьома головними діагоналями розташовані симетрично та кількість їх не залежить від розмірності матриці. Така модифікація має давати точний розв'язок СЛАР та працювати швидше за модифікований метод Гауса.

VII. ВИСНОВКИ

Для рівняння Річардса-Клюта на графі було проаналізовано структуру матриці СЛАР, що виникає на кожному кроці по часу при використанні чисельних методів для знаходження наближеного розв'язку рівняння. Ця матриця має властивості, що дозволяють використовувати для розв'язання СЛАР швидкі чисельні методи Зейделя та спряжених градієнтів, модифіковані для розріджених матриць. Також структура матриці дозволяє використовувати модифікований для розріджених матриць метод Гауса, що дозволяє отримувати точний розв'язок СЛАР і для якого також була наведена оцінка часу його роботи, а саме лінійна залежність часу від розмірності матриці СЛАР з коефіцієнтом пропорційності, рівним кількості вершин графу, в яких повинне виконуватися рівняння балансу маси.

ЛІТЕРАТУРА

- [1] M. Celia, E. Bouloutas, and R. Zabra, "A general mass-conservative numerical solution for the unsaturated flow equation," *Water Resources Research*, vol. 26, no. 1, pp. 1483–1496, Jul. 1990, <https://doi.org/10.1029/WR026i007p01483>
- [2] H. Suk, and E. Park, "Numerical solution of the Kirchhoff-transformed Richards equation for simulating variably saturated flow in heterogeneous layered porous media," *Journal of Hydrology*, vol. 519, no. 124213, Dec. 2019, <https://doi.org/10.1016/j.jhydrol.2019.124213>
- [3] F. Liu, Y. Fukumoto, and X. Zhao, "A multi level linearized Crank–Nicolson scheme for Richards equation under variable flux boundary conditions," *Applicable Analysis*, vol. 102, no. 6, pp. 1601–1617, Oct. 2023, <https://doi.org/10.1080/00036811.2021.1992395>
- [4] G.J. Machado *et al.*, "A new stabilized scheme for a Richards' equation with evapotranspiration," *Groundwater for Sustainable Development*, vol. 17, no. 100736, May 2022, <https://doi.org/10.1016/j.gsd.2022.100736>
- [5] M.W. Farthing, and F.L. Ogden, "Numerical Solution of Richards' Equation: A Review of Advances and Challenges," *Soil Science Society of America*, vol. 81, no. 6, pp. 1257–1269, Dec. 2017, <https://doi.org/10.2136/sssaj2017.02.0058>
- [6] Y. Zha *et al.*, "Zha Y., Yang J., Zeng J., Tso C.-H.M., Zeng W., Shi L. Review of numerical solution of Richardson–Richards equation for variably saturated flow in soils," *WIREs Water*, vol. 6, pp. e1364, Jun. 2019, <https://doi.org/10.1002/wat2.1364>

Розробка автоматизованої системи розрахунку та прогнозування показників екологічного стану

Аліна Ольвінська, Вікторія Рувінська

*Кафедра інженерії програмного забезпечення
Національний університет «Одеська політехніка»
Одеса, Україна*

Анотація — Визначено основні вимоги до програмного продукту на основі аналізу наявних рішень в екологічній сфері. Сформовано основний функціонал програмного продукту та обґрунтовано методики та моделі, що будуть використовуватися для розрахунку та прогнозування показників екологічного стану.

Ключові слова — *Інженерія програмного забезпечення, Екологія, Long short-term memory networks, Автоматизована система.*

I. Вступ

У наш час питання розвитку в екологічному секторі є актуальним, через проблеми, що впливають з поганого екологічного стану в окремому регіоні, країні та загалом у світі. Подальший розвиток та покращення екологічного стану в Україні вже зараз потребує створення інструментів та методологій для спрощення та підвищення ефективності процесів, що необхідні для прийняття більш екологічних рішень при подальшому розвитку країни. Це потребує ознайомлення з показниками екологічного стану та впливу кожного з них на загальний екологічний стан країни. Однак розрахунок та подальший аналіз цих показників потребує роботи з великим масивом даних, особливо, коли аналіз необхідно проводити на основі інформації за кілька років та з декількох регіонів. Прийняття рішень у сфері екології також може потребувати прогнозування впливу змін у показниках на зміну загального екологічного стану, що також потребує роботи з великим набором інформації. Це, звісно, приводить до великих витрат часу, ресурсів та збільшення вірогідності помилок. Для пришвидшення розрахунків та отримання даних у багатьох сферах можна використовувати допомогу автоматизованих систем. Такі системи допомагають полегшити такі процеси як розрахунки або аналіз великих масивів даних. До того ж автоматичні розрахунки значно підвищують не тільки ефективність, але й надійність та доступність розрахунку для користувачів.

Однак наявні програмні рішення, що найчастіше є набором різних статистичних методологій, потребують спеціальних знань та навичок, що зменшує доступність даних рішень та змушує користувача витратити час на пошук необхідного методу та ознайомлення з ним. На меті розробка програмного продукту, який надавав би користувачу можливість розрахувати, спрогнозувати та аналізувати показники екологічного стану заданого регіону на основі введеної або завантаженої інформації за допомогою заздалегідь налаштованих методів та моделей, а також надавав користувачу необхідну інформацію про показники та їх вплив на загальний екологічний стан у зрозумілому графічному вигляді.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Для розуміння вимог до програмного продукту, що розробляється, було проведено аналіз наявних рішень у сфері екологічного аналізу та прогнозування. З боку аналізу та розрахунків було виділено такі загальні програми для статистичного аналізу як Genstat та Canoco та програму для багатовимірної аналізу екологічних даних PC-ORD. Програми для статистичного аналізу надають користувачу набір статистичних методів для дослідження завантаженої інформації та мають інструменти для візуалізації результатів. PC-ORD також надає різні методи для статистичного аналізу, але з фокусом саме на екологічну сферу. З боку прогнозування маємо такі програмні рішення як Minitab та STATGRAPHICS. Обидва рішення мають великий набір методів як для аналізу, так і для прогнозування. Дані програмні рішення також надають можливість візуалізувати отримані прогнози. Таким чином вже існують рішення для автоматизованого статистичного аналізу та, навіть, прогнозування, але усі вони потребують саме знань та навичок зі статистики. З іншого боку, робота з екологічними даними також вимагає й екологічних знань. Даний поріг знань можна знизити, якщо надати користувачу зменшений, але зосереджений на заданому завданні, функціонал та надати необхідну інформацію про нього усередині самої програми.

III. МЕТОДИ

Отже, як результат аналізу аналогів системи, що розробляється, можемо зробити висновок, що наша система повинна бути доступною та зрозумілою для користувачів, мати ефективний, але інтуїтивно зрозумілий функціонал та зменшити вплив людського фактора на процес розрахунку завдяки наданню їм вже готових функцій та моделей для розрахунку, аналізу та прогнозування показників екологічного стану. Таким чином маємо надати користувачу такий функціонал:

- завантаження даних – система повинна отримати файл від користувача та зчитати інформацію з нього до пам'яті програми для подальшого використання;
- розрахунок показників – система повинна розрахувати показники екологічного стану та демонструвати результати користувачу;
- прогнозування показників - система повинна прогнозувати динаміку показників екологічного стану та демонструвати користувачу результати.
- аналіз показників – система повинна аналізувати які чинники найбільш негативно вплинули на показники повітря, води та ґрунту, а також на загальний екологічний стан.
- збереження даних – збереження у файл попередньо отриманих даних у зрозумілому та стандартизованому форматі.

У рамках цієї розробки буде використовуватися інтегральний показник екологічного стану регіонів, який являє собою середньгеометричну з окремих коефіцієнтів, що характеризують стан окремих елементів навколишнього середовища, а саме водних ресурсів, атмосферного повітря, земельних та лісових ресурсів:

$$I = \sqrt[n]{k_1 * k_2 * k_3 \dots * k_n},$$

де k - нормований показник, що характеризує окремі елементи навколишнього середовища.

Оскільки показники, що характеризують окремі елементи навколишнього середовища дуже різноманітні, необхідно провести нормування за формулами:

$$\text{для показників стимуляторів} - k = \frac{x}{x_{max}},$$

$$\text{для показників дестимуляторів} - k = \frac{x_{min}}{x},$$

де x_{max} – максимальне значення показника серед областей України,

x_{min} – мінімальне значення показника серед областей України,

Розрахунок даного показника було зроблено за методикою побудови інтегральних показників [1, р. 8]. На основі даних розрахунків можна проаналізувати вплив показників на загальний екологічний стан регіону.

Для прогнозування буде використовуватися модель машинного навчання. Аналіз наявних методів показав, що у прогнозуванні екологічних процесів та впливу на стан навколишнього середовища вже використовувалися такі моделі як Random Forest Regression та Long Short-Term Memory networks (LSTM) [2]-[4]. В основі Random Forest лежать дерева прийняття рішень, які будуються за допомогою машинного навчання, але вони або не враховують фактор часу або цей фактор опосередковано. Для екологічних моделей час є суттєвою характеристикою, через те, що екологічні показники мають послідовність у часі, тому такого роду дані доцільно представляти за допомогою часових рядів як послідовності значень. LSTM переважно використовуються для вивчення, обробки та класифікації послідовних даних, оскільки ці мережі можуть вивчати довгострокові залежності між часовими кроками даних. Через присутність багатьох пов'язаних між собою показників, кожен з яких впливає на загальний стан регіону, LSTM краще підійде для прогнозування змін у якості екологічного стану в залежності від змін у відповідних показниках у деякому проміжку часу.

IV. РЕЗУЛЬТАТИ

На даному етапі було визначено основні функціональні та нефункціональні вимоги до програмного продукту, а також обрано методології для розрахунку та прогнозування показників екологічного стану.

V. ОБГОВОРЕННЯ

Запропоноване програмне рішення може сприяти спрощенню та пришвидшенню процесу розрахунку та аналізу показників екологічного стану та оцінки їх впливу на екологічний стан окремих регіонів та країни загалом, а можливість прогнозувати ці показники на майбутнє допоможе у розробці регіональних програм охорони навколишнього середовища та прийняття обґрунтованих управлінських рішень як на регіональному, так і державному рівні.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Подальший розвиток запропонованого програмного рішення потребує розробки прототипу продукту, його тестування та подальшого покращення на основі відгуків тестувальників.

VII. ВИСНОВКИ

Визначено основні вимоги до програмного продукту на основі аналізу наявних рішень. Сформовано основний функціонал програмного продукту та обґрунтовано методики та моделі, що будуть використовуватися для розрахунку та прогнозування показників екологічного стану.

ЛІТЕРАТУРА

- [1] Sydorova A. V., Bilenko D. V., Burkina N. V. *Biznes-analytika: navchal'no-metodychnyy posibnyk*. [Business analytics: educational and methodological guide] Vinnytsya: DonNU imeni Vasylyya Stusa. 2019.
- [2] M. Khosravi, B. M. Duti, M. M. S. Yazdan, S. Ghoochani, N. Nazemi, and H. Shabaniyan, "Multivariate Multi-Step Long Short-Term Memory Neural Network for Simultaneous Stream-Water Variable Prediction," *Eng*, vol. 4, no. 3, pp. 1933–1950, Jul. 2023, doi: 10.3390/eng4030109. [Online]. Available: <http://dx.doi.org/10.3390/eng4030109>
- [3] Haq, M.A., Ahmed, A., Khan, I. et al. Analysis of environmental factors using AI and ML methods. *Sci Rep* 12, 13267 (2022). [Online]. Available: <https://doi.org/10.1038/s41598-022-16665-7>
- [4] A. Koyampambath, N. Adibi, C. Szablewski, S. A. Adibi, and G. Sonnemann, "Implementing Artificial Intelligence Techniques to Predict Environmental Impacts: Case of Construction Products," *Sustainability*, vol. 14, no. 6, p. 3699, Mar. 2022, doi: 10.3390/su14063699. [Online]. Available: <http://dx.doi.org/10.3390/su14063699>

Розробка системи розділення джерел звуку на основі методів машинного навчання

Орест Ткачук

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація — Ця стаття націлена на розгляд можливостей використання згорткових нейронних мереж (CNN) для розділення акустичних джерел у аудіосигналах. Дослідження базується на аналізі особливостей, які можуть бути виділені завдяки використанню віконного перетворення Фур'є, що надає можливість аналізувати аудіосигнали в часово-частотному домені і розкриває їх часові та частотні характеристики. Додатково, розглядається можливість використання бінарних масок як інструменту для спрощення завдання розділення акустичних джерел. Цей підхід потенційно може покращити існуючі методи розділення акустичних джерел, розширити можливості обробки аудіосигналів у різних застосуваннях та зробити їх більш ефективними та потужними. Детальний аналіз та практичні дослідження розв'язання цих питань можуть допомогти вдосконалити наявні методи та визначити нові перспективи в цій області.

Ключові слова— *Комп'ютерні науки, Штучний інтелект, Згорткові нейронні мережі, Аналіз аудіо даних, Обробка аудіосигналів, Розділення аудіо джерел.*

I. ВСТУП

Сучасні технології в обробці звуку дозволяють досягти високої якості аудіо та звукових систем, необхідної для багатьох галузей, таких як медіа, телекомунікації та безпека. Однак однією з важливих викликів в цій сфері є потреба в розділенні звукових джерел на окремі компоненти. Наприклад, у музичній обробці чи розпізнаванні мовлення часто необхідно виділити окремі акустичні джерела, які накладаються одне на одне у вихідних аудіофайлах.

Методи розділення звукових джерел, знаходять широке застосування в різних галузях. Наприклад, в музичній індустрії вони дозволяють автоматично виділити окремі інструментальні компоненти аудіозаписів для їхнього редагування та міксування. У галузі безпеки, ці методи використовуються для ідентифікації голосів та виділення аудіосигналів, що може бути корисним у розслідуванні. Крім того, в медицині вони застосовуються для виділення конкретних звукових сигналів, які використовуються в діагностиці та моніторингу пацієнтів, таких як аудіозаписи серцевих ритмів або дихальних функцій. Ці методи допомагають поліпшити якість та точність обробки аудіо та звуку у різних важливих контекстах.

Для вирішення цього завдання, важливого для багатьох сфер, використовуються методи машинного навчання. Зокрема, глибоке навчання та нейронні мережі дозволяють досягнути високої точності у розділенні звукових джерел, розкриваючи потенціал для подальшого вдосконалення систем обробки аудіо та звуку.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Розділення джерел звуку – актуальна проблема в обробці звуку. Вона полягає в виділенні аудіо-джерел зі змішаного матеріалу. Це завдання варіюється в залежності від контексту, включаючи виділення голосів, шуму, музичних інструментів тощо.

Існують різні методи та алгоритми для розв'язання цієї задачі, проте вона залишається складною через унікальні варіації у завданнях. У цьому дослідженні вивчено кілька важливих статей, які мають взаємозв'язок із даною роботою. Автори [1] розглядали використання невід'ємної матричної факторизації (NMF) для розділення аудіоджерел. Вони запропонували нову модель NMF, яка автоматично визначає кількість джерел в сигналі. Однак їх робота не вирішує всі технічні проблеми та не підходить для всіх типів даних. У статті [2] описано нові підходи до розділення мовленнєвих сигналів за допомогою моделі NMF, що показали покращення якості. Однак ці методи обмежені у разі схожих ознак різних джерел. Автори [3] пропонують напівнавчання, використовуючи набір DNN для розв'язання проблеми розпізнавання мовлення в умовах різного рівня сигнал-шум. Вони досягли кращих результатів, але не враховують всі аспекти змішування мовлення. У роботі [4] розглянуто розділення аудіосигналів на мову, музику та шум. Модель MTASSNet ефективна, але є обмеження для складних міксів.

Багато підходів в розділенні джерел звуку використовують стандартну невід'ємну матричну факторизацію, але ці методи мають обмеження, особливо у разі схожих ознак різних джерел. Останнім часом стали популярними підходи з використанням нейронних мереж. Їх використання може допомогти вирішити деякі проблеми, що виникають при використанні NMF, проте цей підхід вимагає додаткового дослідження гіперпараметрів. У цій статті досліджуватиметься розширене використання нейронних мереж (NN) для вирішення задачі розділення джерел звуку, з основним акцентом на використанні згорткових нейронних мереж (CNN) як потенційного методу для покращення результатів цієї задачі.

III. МЕТОДИ

У цій роботі використовується віконне перетворення Фур'є (STFT), яке є ефективним інструментом для представлення аудіосигналів у часово-частотному домені. STFT дозволяє аналізувати частотні характеристики звуку в різні моменти часу. Це важливо, оскільки різні акустичні джерела можуть мати різні часово-частотні особливості, які дозволяють їхнє виділення та розділення (рис.1). Структура STFT має формулу, що дозволяє представляти звукові сигнали через часові вікна та їх частотні складові. Віконне перетворення виконується відповідно до виразу:

$$F(\omega, b) = \int_{-\infty}^{+\infty} f(x)w(x - b)e^{-i\omega x} dx$$

Функція $w(x - b)$ являє собою функцію (у загальному випадку комплексну) вікна зсуву перетворення за координатою x , де параметром b задаються фіксовані значення зсуву.

Крім того, згорткові нейронні мережі використовуються для обробки спектрограм, отриманих в результаті STFT. Це дозволяє CNN виявляти акустичні шаблони і властивості у вигляді зображень, що відповідає різним часовим та частотним особливостям звукових сигналів. Комбінування STFT з CNN відкриває нові можливості для визначення та розділення різних джерел звуку в аудіоматеріалах.

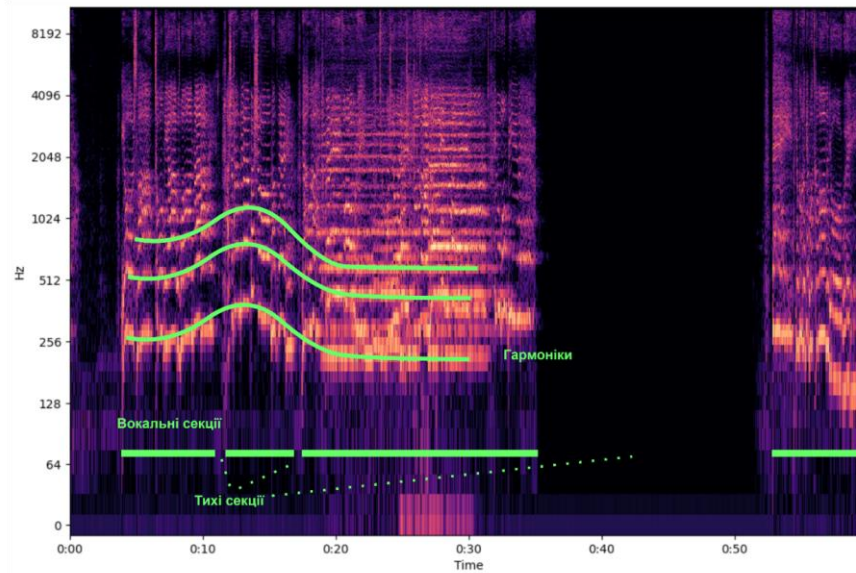


Рис.1. Спектрограма вокалу з позначеннями акустичних особливостей

Для ефективної реконструкції акустичних патернів, зображених на спектрограмі STFT та оброблених за допомогою CNN, важливо враховувати, що ця задача може виявитися складною для нейронної мережі. У зв'язку з цим можна розглянути альтернативний підхід, який полягає в визначенні бінарних масок. Замість спроби безпосередньо реконструювати STFT для конкретного джерела, визначається бінарна маска, яка показує наявність (1) або відсутність (0) кожного джерела звуку для кожного часового кадру та частотного піддіапазону на вхідній спектрограмі. Ця бінарна маска (В) розраховується за допомогою формули:

$$B = (S > T * M),$$

де В – бінарна маска, S – інтенсивність сигналу відокремленого джерела (наприклад, вокалу) на вхідній спектрограмі, а М – інтенсивність сигналу суміші аудіосигналів на вхідній спектрограмі.

Ця бінарна маска використовується для виділення конкретних джерел у вихідному міксі, що допомагає покращити розділення джерел звуку.

За допомогою CNN, модель може бути навчена розглядати не лише окремі вікна мікса для передбачення конкретного інструменту, а використовувати 2-вимірні зображення, отримані з набору вікон. Цей підхід дозволяє зберігати темпоральний контекст, що покращує ефективність аналізу (рис.2.). Крім того, для випадків, коли неможливо взяти вікна до або після, такі як на межі запису, можна продублювати вже наявні крайні вікна, забезпечуючи збереження коректності передбачень.

Загальна архітектура використаної CNN зображена на рис.3. Дана архітектура мережі включає два блоки з двома шарами згортки і одним шаром макс-пулінгу в кожному. Після макс-пулінгу застосовуються дропаут зі значенням 0.1 для регуляризації. Далі дані перетворюються в одновимірний формат, і до них застосовується повнозв'язний шар. Потім використовується ще один дропаут зі значенням 0.2 для подальшої регуляризації, і завершальний повнозв'язний шар генерує вектор із 513 значень в потрібному форматі.

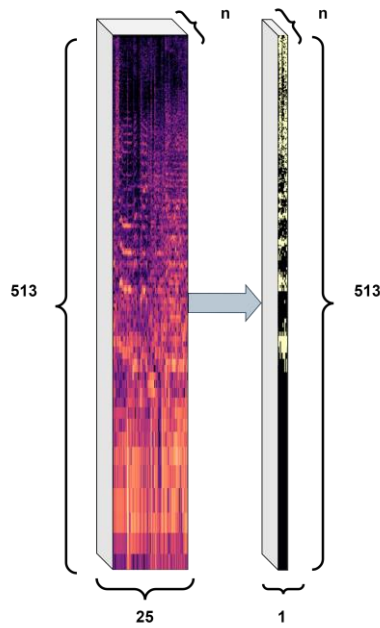


Рис.2. Схема передбачення бінарної маски джерела

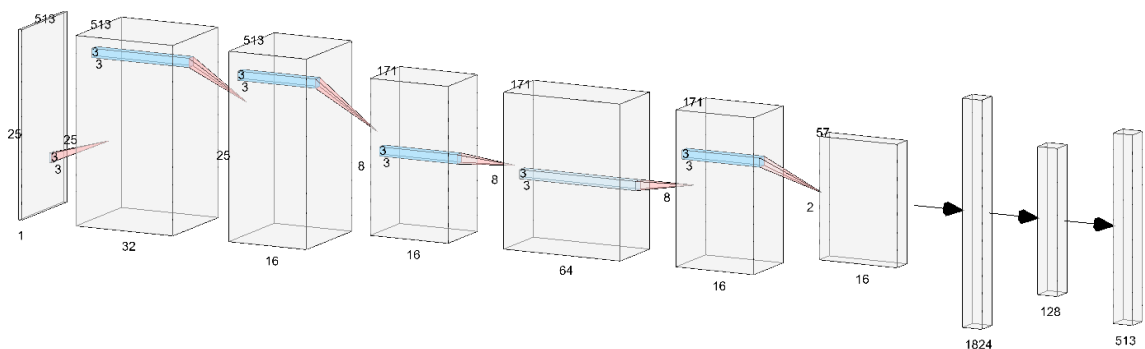


Рис.3. Архітектура CNN для розділення джерел

Набір даних MUSDB18 [5], [6], використаний у даній статті, є збіркою аудіофайлів, які представлені у форматі stem.mp4. Ці файли містять набір аудіо стемів, що репрезентують окремі звукові доріжки з оригінальних музичних композицій.

Зазначений набір даних складається з 100 тренувальних та 50 тестових композицій з частотою дискретизації 44100 Гц. Кожна композиція розділена на окремі аудіо-доріжки для різних компонентів музичних композицій, таких як інструменти та вокал. Це дозволяє окремо аналізувати та маніпулювати окремими аспектами звукових записів під час обробки аудіосигналів, як інструментальні частини чи голосові треки. Для роботи з цим набором даних потрібно використовувати віконне перетворення Фур'є, щоб отримати амплітудно-часове представлення аудіосигналів у вигляді спектрограм. Це дозволить аналізувати зміни амплітуди сигналів відносно часу та частоти, що є важливим для подальшої обробки та аналізу аудіоданих.

У даній задачі передбачення бінарної маски використовується гібридний підхід, що поєднує елементи регресії та класифікації. Модель пробує класифікувати пікселі на виході, щодо того, чи вони віднесені до конкретного джерела чи ні. Оскільки концептуально

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

завдання лишається регресійним, функція втрат визначається як Mean Squared Error (MSE). Для вимірювання точності результати, отримані від згорткової нейронної мережі, округлюються, і перевіряється, яка частина з них відповідає реальним даним, а яка – ні.

IV. РЕЗУЛЬТАТИ

Розглянемо конкретні числові параметри, які використовуються у даній роботі. Частота дискретизації (fs) становить 22050 Гц, після зниження її з рівня 44100 Гц. Для віконного перетворення Фур'є ми використовуємо розмір вікна 1024 і крок 256, що призводить до створення 513 частотних діапазонів. Роздільна здатність одного вікна STFT становить приблизно 11.6 мілісекунд. Для вирішення завдання ми використовуємо часовий контекст тривалістю приблизно 300 мілісекунд, що відповідає 25 вікнам STFT. Планована кількість навчальних прикладів складає 100 пісень по 60 секунд кожна, що становить 100 пісень по близько 5000 вікон STFT, або близько 0,5 мільйона зразків даних. Для позначення і створення цієї кількості даних при використанні ковзного вікна з кроком 1 в часовому проміжку STFT знадобиться приблизно 1.6 години аудіоматеріалу (60 секунд на кожному з 100 пісень).

Використовуючи описану модель згорткової нейронної мережі, було проведено навчання на окремих тренувальних наборах даних для кожного джерела звуку. Процес навчання тривав 50 епох і включав в себе використання стохастичного градієнтного спуску (SGD) як оптимізатора та циклічного планування обчислювальних швидкостей (CyclicLR) зі значеннями, що коливалися в діапазоні від 0.1 до 0.4. Для додаткової регуляризації моделей також було використано шари викидів з частками відключень нейронів, а саме 0.1, 0.1 та 0.2.

Таб.1. Результати точності та функції втрат для кожного джерела

Тип джерела для ізоляції	Тип набору даних	Точність	Результат функції втрат(MSE)
<i>Ізоляція вокалу</i>	Тренувальний	0.772	0.153
	Валідаційний	0.743	0.174
<i>Ізоляція барабанів</i>	Тренувальний	0.766	0.159
	Валідаційний	0.685	0.203
<i>Ізоляція басів</i>	Тренувальний	0.944	0.044
	Валідаційний	0.939	0.051
<i>Ізоляція решти джерел</i>	Тренувальний	0.764	0.158
	Валідаційний	0.732	0.1791

У зв'язку з тим, що загальний музичний твір представляє собою складну комбінацію різних джерел звуку, можна констатувати, що загальна точність у відокремленні цих джерел може бути оцінена як середнє значення точності виділення кожного окремого джерела.

Таб.2. Усереднені результати точності та функції втрат

<i>Тип набору даних</i>	Точність	Результат функції втрат (MSE)
<i>Тренувальний</i>	0.811	0.129
<i>Валідаційний</i>	0.775	0.152

З отриманих результатів можна побачити що дана архітектура CNN добре справилася з знаходженням особливостей в бінарних масках джерел, і здатна виділяти окремі джерела з загального музичного твору.

V. ОБГОВОРЕННЯ

Результати, представлені в таблицях 1 і 2, розкривають важливі аспекти дослідження щодо ізоляції різних джерел аудіосигналів. Варто відзначити декілька ключових пунктів.

Результати вказують на те, що точність моделі варіюється в залежності від типу ізолюваного аудіоджерела. Найвищу точність в 94% отримано при ізоляції басів як на тренувальному, так і на валідаційному наборах даних. Це свідчить про високу ефективність моделі у виділенні цього аудіокомпонента.

Цікавим аспектом є той факт, що, незважаючи на відносно низьку точність при ізоляції барабанів(76.6% на тренувальному і 68.5% на валідаційному), аудіозапис, який був ізолюваний з цього джерела, має досить чистий звук. Схожий ефект спостерігається при ізоляції басів, де звук також залишається досить чистим, проте на відміну від барабанів точність там значно вища. Це може бути пов'язано з низьким рівнем перекриття між різними аудіокомпонентами, що полегшує їх ідентифікацію.

Загальні результати, усереднені для тренувального та валідаційного наборів даних становлять 81% та 77% відповідно і демонструють, що дана модель, незважаючи на простоту архітектури, успішно вирішує поставлену задачу. Варто зазначити, що основним обмеженням роботи є саме проста архітектура моделі, і є потенціал для подальшої оптимізації та покращення результатів.

У цілому, результати дослідження вказують на потенціал такого підходу у вирішенні задачі ізоляції аудіосигналів різних джерел. Подальші дослідження та розвиток цього напрямку можуть призвести до покращення та оптимізації результатів у майбутньому

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У майбутніх дослідженнях можливо розглянути розширення та оптимізацію архітектури згорткової нейронної мережі. У поточній роботі використовувалася відносно проста модель з обмеженою кількістю шарів та параметрів для компактності і ефективності тренування. Збільшення глибини та додавання додаткових шарів може допомогти нейронній мережі виявляти ще більше важливих особливостей у вихідних даних.

Крім того, варто розглянути альтернативні методи створення бінарної маски. Поточна імплементація надає високу деталізацію, проте може ускладнити задачу передбачення для нейронної мережі. Можливо, варто розглянути підхід, який жертвує деякими деталями, які можуть здаватися схожими на шум, на користь більш ефективного опису загальної структури джерела звуку.

Також варто вивчити можливість розробки архітектури згорткової нейронної мережі, яка працюватиме з уявною частиною віконного перетворення Фур'є. Навіть якщо ця частина зазвичай не розглядається в обробці аудіосигналів, вона все ще містить корисну інформацію, яка може покращити процес виділення звуку та джерела.

VII. ВИСНОВКИ

У даній роботі була поставлена задача ізоляції різних джерел аудіосигналів, зокрема, вокалу, барабанів, басів та інших компонентів. Гіпотеза полягала в тому, що застосування згорткових нейронних мереж дозволить ефективно визначати наявність або відсутність кожного з джерел звуку на вхідній спектрограмі та відокремлювати їх.

Для вирішення цієї задачі було використане віконне перетворення Фур'є, яке дозволяло представляти аудіосигнали у часово-частотному домені. Потім згорткові нейронні мережі використовувалися для обробки отриманих спектрограм. Використання бінарних масок замість прямого STFT дозволило спростити задачу і змінити підхід на гібридний, що поєднує елементи регресії та класифікації. Використання CNN дозволяло моделі виявляти акустичні шаблони та властивості на вхідних спектрограмах, що відповідали різним джерелам звуку.

Результати показали високі показники точності для ізоляції різних джерел аудіосигналу. Зокрема, точність для ізоляції вокалу складала близько 77.1%, для басів – близько 94.1%, для барабанів – близько 76.6%, і для інших джерел – близько 76.4%. Крім того, середньоквадратична помилка також була досить низькою на тренувальному і валідаційному наборах даних, і становила в середньому близько 0.13 для тренувального і 0.16 для валідаційного наборів.

ЛІТЕРАТУРА

- [1] V. Leplat, N. Gillis, and A. M. S. Ang, "Blind Audio Source Separation With Minimum-Volume Beta-Divergence NMF," *IEEE Trans. Signal Process.*, vol. 68, pp. 3400–3410, 2020, doi: 10.1109/TSP.2020.2991801.
- [2] Y. Feng, "Single and Multichannel Speech Source Separation using Non- Negative Matrix Factorisation Incorporating Spectral Masks," *University of Wollongong Thesis Collection 2017*, Jan. 2017, [Online]. Available: <https://ro.uow.edu.au/theses1/90>
- [3] J. Du, Y. Tu, L.-R. Dai, and C.-H. Lee, "A Regression Approach to Single-Channel Speech Separation Via High-Resolution Deep Neural Networks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 8, pp. 1424–1437, Aug. 2016, doi: 10.1109/TASLP.2016.2558822.
- [4] L. Zhang, C. Li, F. Deng, and X. Wang, "Multi-Task Audio Source Separation," in *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Cartagena, Colombia: IEEE, Dec. 2021, pp. 671–678. doi: 10.1109/ASRU51503.2021.9687922.
- [5] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner, "MUSDB18 - a corpus for music separation." Zenodo, Dec. 17, 2017. doi: 10.5281/ZENODO.1117372.
- [6] "MUSDB18 | SigSep." Accessed: Nov. 02, 2023. [Online]. Available: <https://sigsep.github.io/datasets/musdb.html#musdb18-compressed-stems>

Simulation of critical phenomena in the island systems in general relativity with flows

Hennady Shapovalov, Anatoly Kazakov, Vyacheslav Oleynyk

*Odesa Polytechnic National University,
Odesa National University, Ukraine*

Abstract

Mathematical simulation of critical phenomena in island systems was carried out with the aim of predicting the occurrence of bifurcation phases and conditions for the simultaneous coexistence of several phases in particle-like systems within the framework of general relativity. Based on the generalization of the theory of Landau phase transitions and the provisions of Thom's theory of catastrophes, the fulfillment of the conditions of stability, singularity, and the emergence of phase coexistence spaces for the family of solutions of the Plebański and Demiański equations of Einstein-Maxwell, which is a generalization of the Schwarzschild problem, was investigated. In this work, we discuss the properties of electrovacuum island models with local curvature flows. An expression for the Lagrangian of the considered systems is obtained for the purpose of further simulation of critical phenomena taking into account the occurrence of space-time rotations and accelerated motion of particle-like systems.

Keywords— mathematical modeling, phase transitions, phase coexistence, critical phenomena.

I. INTRODUCTION

Mathematical simulation of critical phenomena in systems that are described by island models is relevant. This is due to the fact that island models can be used to predict unstable states of elementary particles, which expands the possibilities of predicting processes in modern nuclear power. The properties of the Plebanski-Demianski [1] island model were studied [2] using a non-orthogonal gradient tetrad. The Plebanski-Demianski model (eight-parameter solution of vacuum Einstein-Maxwell equations) was studied as a model of Schwarzschild, which contains a local singularity.

II. RELATED WORK

Despite the large number of works related to modeling the processes leading to the emergence and decay of elementary particles within the framework of general relativity, there are significant difficulties in predicting the behavior of island systems to which they belong. This is due to the emergence of singularities in solutions to equations or solutions containing complex quantities [1, 2], which reduces practical interest. In this regard, there is interest in alternative approaches that make it possible to circumvent the problems of singularity and the appearance of complex quantities in solutions [3, 8]. Such an approach could be modeling the phase states of space-time and critical phenomena, which can be compared with the processes of emergence and decay of elementary particles [4, 5, 15].

III. METHODOLOGY

In order to study the phase states and critical phenomena of the island system, the well-known Plebański-Demiański metric in real $x^i = (p, q, \sigma, \tau)$ coordinates was presented in the form:

$$ds^2 = \frac{1}{(p+q)^2} \left\{ \frac{1+(pq)^2}{P} dp^2 + \frac{P}{1+(pq)^2} (d\sigma + q^2 d\tau)^2 + \frac{1+(pq)^2}{Q} dq^2 - \frac{Q}{1+(pq)^2} (d\tau - p^2 d\sigma)^2 \right\}, \quad (1)$$

where the real functions of the coordinates are $P = P(p)$, $Q = Q(q)$ and the signature has the form (+, +, +, -) based on the gradient vectors:

$$m_i = \frac{\partial x^0}{\partial x^i} = (1,0,0,0), \quad n_i = \frac{\partial x^1}{\partial x^i} = (0,1,0,0), \quad p_i = \frac{\partial x^2}{\partial x^i} = (0,0,1,0), \quad s_i = \frac{\partial x^3}{\partial x^i} = (0,0,0,1) \quad (2)$$

According to the solution of Plebański and Demiański [1 – 3], the energy-momentum tensor was written in the form:

$$T^{ij} = \begin{pmatrix} -(\alpha + \beta) \frac{(p+q)^4}{(1+(pq)^2)^2} & 0 & 0 & 0 \\ 0 & -(\alpha + \beta) \frac{(p+q)^4}{(1+(pq)^2)^2} & 0 & 0 \\ 0 & 0 & (\alpha + \beta) \frac{(p+q)^4}{(1+(pq)^2)^2} & 0 \\ 0 & 0 & 0 & (\alpha + \beta) \frac{(p+q)^4}{(1+(pq)^2)^2} \end{pmatrix}, \quad (3)$$

where α , β are constants, and

$$\begin{aligned} P(p) &= (-\lambda/6 - \alpha) + 2np - \varepsilon p^2 + 2mp^3 + (-\lambda/6 - \beta)p^4 \\ Q(q) &= (-\lambda/6 - \alpha) + 2nq - \varepsilon q^2 + 2mq^3 + (-\lambda/6 - \beta)q^4 \end{aligned}, \quad (4)$$

and λ , n , ε , m are constants.

To determine the spaces in which the stability conditions are met on the phase diagram of the T^{ij} components, a system of equations and inequalities was solved [4-12]:

$$\frac{\partial T^{ij}}{\partial p} = 0; \quad \frac{\partial T^{ij}}{\partial q} = 0; \quad \det \frac{d^2 T^{ij}}{dX^2} > 0, \quad (5)$$

where the expressions $\frac{\partial T^{ij}}{\partial p}$, $\frac{\partial T^{ij}}{\partial q}$ denote the partial derivatives of the T^{ij} of the energy-momentum tensor with respect to independent p and q coordinates. To calculate the zero contours of the first partial derivative components x and y , analytical expressions were found, which were presented in the form:

$$\frac{\partial T^{11}}{\partial p} = \frac{\partial T^{22}}{\partial p} = 4 \frac{CB^3}{A^2} \left(1 - \frac{q^2 Bp}{A}\right), \quad \frac{\partial T^{11}}{\partial q} = \frac{\partial T^{22}}{\partial q} = 4 \frac{CB^3}{A^2} \left(1 - \frac{p^2 Bq}{A}\right), \quad (6)$$

where $C = (-\beta - \alpha)$, $A = p^2 q^2 + 1$, $B = q + p$.

In system (5), the expression $\det \frac{d^2 T}{dX^2}$ denotes the determinant of the matrix of the second partial derivative components of the energy-momentum tensor T^{ij} along the coordinates p and q , that is, along $X(p, q)$. To solve system (5), its analytical form was found and presented in the form:

$$\det \frac{d^2 T^{11}}{dX^2} = \det \frac{d^2 T^{22}}{dX^2} = \frac{16C^2 B^4}{A^4} \left[3 - \frac{p^2}{A} B(B+8q) + \frac{6p^4 q^2 B^2}{A^2} \right] \left[3 - \frac{q^2}{A} B(B+8p) + \frac{6p^2 q^4 B^2}{A^2} \right] - \frac{144C^2 B^4}{A^4} \left[1 - \frac{2pq}{A} \left(1 - \frac{p^2 q^2 B^2}{A} \right) \right]^2 \quad (7)$$

$$\det \frac{d^2 T^{33}}{dX^2} = \det \frac{d^2 T^{44}}{dX^2} = -\det \frac{d^2 T^{11}}{dX^2}$$

The matrix of partial derivatives of the momentum energy tensor used in (5) for the analysis of the space signature of the phase diagram of the existence of the island system corresponding to (1) – (3) was constructed according to the rule:

$$\frac{d^2 T^{ij}}{dX^2} = \begin{pmatrix} \frac{\partial^2 T^{ij}(p,q)}{\partial p^2} & \frac{\partial^2 T^{ij}(p,q)}{\partial p \partial q} \\ \frac{\partial^2 T^{ij}(p,q)}{\partial q \partial p} & \frac{\partial^2 T^{ij}(p,q)}{\partial q^2} \end{pmatrix} \quad (8)$$

To determine the spaces in which the conditions for the existence of the bifurcation space are met on the phase diagram of the T^{ij} components, a system of equations was solved by a method similar to the solution of system (5) [4 – 12]:

$$\frac{\partial T^{ij}}{\partial p} = 0; \quad \frac{\partial T^{ij}}{\partial q} = 0; \quad \det \frac{d^2 T^{ij}}{dX^2} = 0. \quad (9)$$

To clarify the fulfillment of the conditions for the occurrence of phase coexistence spaces, a system of equations and inequalities was solved, to which, in addition to the fulfillment of conditions (9), the conditions of zero values of the third partial derivative components of the tensor T^{ij} along the $X(p,q)$ coordinates were added, as well as the condition of a positive signature was added of the determinant of the fourth partial derivatives matrix:

$$\begin{aligned} \frac{\partial T^{ij}}{\partial p} = 0; \quad \frac{\partial T^{ij}}{\partial q} = 0; \quad \det \frac{d^2 T^{ij}}{dX^2} = 0, \\ \frac{\partial^3 T^{ij}}{\partial p^3} = 0, \quad \frac{\partial^3 T^{ij}}{\partial p^2 \partial q} = 0, \quad \frac{\partial^3 T^{ij}}{\partial p \partial q \partial p} = 0, \quad \frac{\partial^3 T^{ij}}{\partial p \partial q^2} = 0, \quad \frac{\partial^3 T^{ij}}{\partial q \partial p^2} = 0, \quad \frac{\partial^3 T^{ij}}{\partial q \partial p \partial q} = 0, \quad \frac{\partial^3 T^{ij}}{\partial q^2 \partial p} = 0, \quad \frac{\partial^3 T^{ij}}{\partial q^3} = 0, \quad (10) \\ \det \frac{d^4 T^{ij}}{dX^4} > 0, \end{aligned}$$

where $\det \frac{d^4 T^{ij}}{dX^4}$ is the determinant of the matrix of the fourth partial derivatives of the components of T^{ij} along the $X(p,q)$ coordinates [14]:

$$\frac{d^4 T^{ij}}{dX^4} = \begin{pmatrix} \left(\frac{\partial^4 T^{ij}(p,q)}{\partial p^4} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial p^3 \partial q} \right) & \left(\frac{\partial^4 T^{ij}(p,q)}{\partial p \partial q \partial p^2} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial p \partial q \partial p \partial q} \right) \\ \left(\frac{\partial^4 T^{ij}(p,q)}{\partial p^2 \partial q \partial p} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial p^2 \partial q^2} \right) & \left(\frac{\partial^4 T^{ij}(p,q)}{\partial p \partial q^2 \partial p} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial p \partial q^3} \right) \\ \left(\frac{\partial^4 T^{ij}(p,q)}{\partial q \partial p^3} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial q \partial p^2 \partial q} \right) & \left(\frac{\partial^4 T^{ij}(p,q)}{\partial q^2 \partial p^2} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial q^2 \partial p \partial q} \right) \\ \left(\frac{\partial^4 T^{ij}(p,q)}{\partial q \partial p \partial q \partial p} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial q \partial p \partial q^2} \right) & \left(\frac{\partial^4 T^{ij}(p,q)}{\partial q^3 \partial p} \quad \frac{\partial^4 T^{ij}(p,q)}{\partial q^4} \right) \end{pmatrix} \quad (11)$$

System (10) defines the conditions when one stable state is replaced by another stable state. This is a first-order transition where the two minima have the same depth. The location of the points of the phase diagram where conditions (10) are fulfilled corresponds to the space of simultaneous coexistence of two phases [5, 15].

IV. RESULTS

Taking into account that α , β are constants, the spaces in which the stable phase condition (5) is fulfilled were determined, as shown in Fig. 1. The zero contours of the determinants (7) and their topology were determined using the open computer algebra system MAXIMA [13] and direct calculations around the found zero contours of the energy-momentum tensor components.

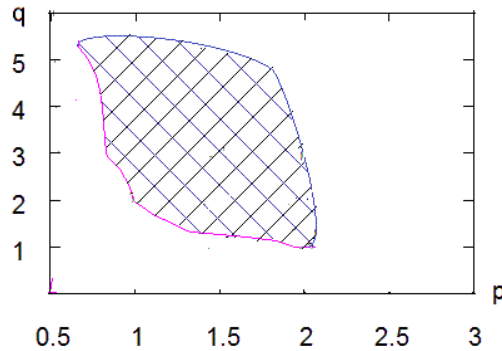


Fig. 1. The phase diagram space in which the stable phase condition is satisfied for the T^{11} and T^{22} components.

The results of solving system (9) are shown in Fig. 2:

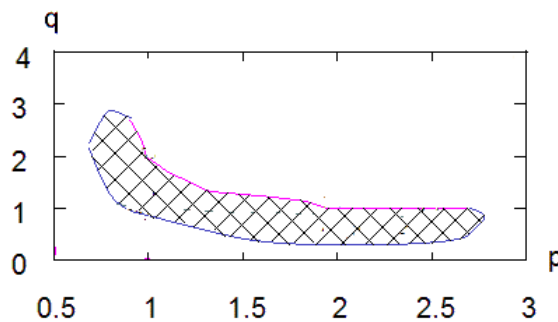


Fig. 2. The phase diagram space in which the bifurcation space condition for T^{11} and T^{22} is satisfied.

The results of solving system (9) are shown in Fig. 2:

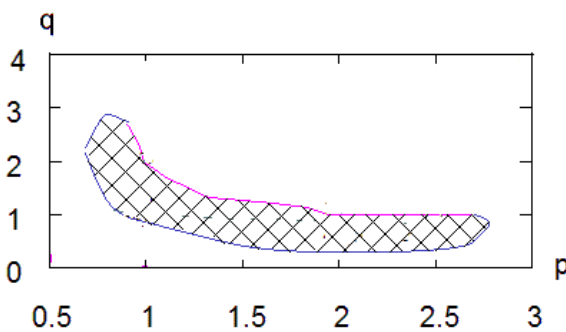


Fig. 2. The phase diagram space in which the bifurcation space condition for T^{11} and T^{22} is satisfied.

V. DISCUSSION

According to the provisions of Thom's theory of catastrophes and the generalization of the theories of Ginsburg-Landau phase transitions using the differential topological approach, it was found that the island system that corresponds to the Plebański-Demiański metric has spaces on the phase diagram of existence in which the conditions for the emergence of a stable phase, bifurcation space and space of coexistence of at least two phases at the same time. Based on the results of calculations of zero contours and the signature of the higher derivative components of the energy-momentum tensor, the positions of the spaces on the phase diagram in the coordinates of the Plebański-Demiański metric, where the conditions for the existence of a stable phase of the system are fulfilled, were obtained. The locations of the set of points of the phase diagram where the conditions for the occurrence of the bifurcation space and the space of phase coexistence of order two are fulfilled are calculated. Based on the obtained results of modeling the positions of spaces of coexistence of phases of order two, it can be assumed that the model presented in the work can be used to study the possibility of the emergence of spaces of coexistence of phases of higher orders. A more detailed analysis of the simulation results as applied to specific cosmological objects is associated with specifying specific values of the constants included in expressions (3) and (4). In work (1), these constants correspond to the mass of the object under study, NUT parameter, angular momentum per unit mass, acceleration, electric and magnetic charge.

VI. FUTURE RESEARCH

However, the results obtained do not take into account the effects that may arise due to the rotational moment in the island system and accelerated motion. To take into account these effects, an expression was obtained for the Lagrangian taking into account rotation and accelerated motion:

$$L_0 = \frac{1}{2DF} \{ Z[D^\Delta(AC^\Delta + 2BB^\Delta + CA^\Delta) + F^*(AC^* + 2BB^* + CA^*) - FB^{*2} - DB^{\Delta 2}] - Z^2[B^2(FA^*C^* + DA^\Delta C^\Delta + 2FB^{*2} + 2DB^{\Delta 2}) + AC(FA^*C^* + DA^\Delta C^\Delta)] + \frac{FD^{*2}}{D^2} + \frac{DF^{\Delta 2}}{F^2} \}$$

The symbols * and Δ mean differentiation with respect to the coordinates p and q , respectively. Functions in metric (1) taking into account the effects of rotation and acceleration have the form:

$$\begin{aligned} P(p) &= -(\lambda/6) + \gamma - \varepsilon p^2 + (-\lambda/6 - \gamma) p^4 \\ Q(q) &= -(\lambda/6) + \gamma + \varepsilon q^2 + (-\lambda/6 - \gamma) q^4 \end{aligned} \quad (5)$$

where

$$\varepsilon = -\frac{a^2 - b^2}{ab(a^2 + b^2)} \left[1 - \frac{\lambda}{3}(a^2 + b^2) \right]^{1/2} \quad \gamma = \frac{1}{a^2 + b^2} - \frac{\lambda}{6} \quad (6)$$

The parameter a is related to the rotation (angular momentum per unit mass) and the parameter b has the meaning of the inverse of the acceleration.

The next stage in the study of the phase states of particle-like systems is carried out in accordance with the obtained expression for the Lagrangian of the system.

REFERENCES

1. Plebanski J.F., Demianski M. Rotating, Charged, and Uniformly Accelerating Mass in General Relativity / Ann. Phys. (NY), 1976, V. 98, P. 98-127.
2. Shapovalov, G.V. 16-th Odessa Int. Astr. Gamow Conf.-School, 2016, P. 12.
3. Shapovlov H. Mathematical modeling of critical phenomena according to the Plebiansky-Demyansky metric / Shapovlov H., Kazakov A., Oleynik V., Zorilo V. // Technologies, Innovative And Modern Theories Of Scientists. – 2023 2nd International Conference on Innovative Solutions in Software Engineering Ivano-Frankivsk, Ukraine, November 29-30, 2023

Proceedings of the XX International Scientific and Practical Conference. – Graz, Austria. – May 23 – 26, 2023, P. 434 – 438

4. Cahn, J. On spinodal decomposition / J.Cahn // *Acta Met.* — 1961. — Vol. 9. — P. 795 — 801.
5. Okada, K. Classical calculations on the phase transition I. Phase diagram in four-dimensional space for the system with one order parameter / K. Okada, I. Suzuki // *J. Phys. Soc. Jap.* — 1982. — Vol. 51, № 10. — P. 3250 — 3257.
6. Laugier, A. Thermodynamics and phase diagram calculations in II-VI and IV-VI ternary systems using an associated solution model / A. Laugier // *Revue De Physique Applique.* — 1973. — Vol. 8, № 9 — P. 259–26.
7. Kazakov, A. Computer simulation for stability of quaternary solid solutions / A. Kazakov, I.Kishmar // *J. Crystal Growth.* — 1991. — Vol. 110. — P. 803 — 814.
8. Kazakov, A. Stability analysis of quaternary alloys including the lattice mismatch strain energy / A.Kazakov, I. Kishmar // *J. Crystal Growth.* — 1992. — Vol. 125. — P. 509 — 518.
9. Kazakov, A. Calculation of the phases coexistence spaces in the system Hg-Mn-Te-Se [Текст] / A. Kazakov, D. Burtnyi, G. Shapovalov // *Proceedings of Odessa Polytechnic University.* — 2018. — T. 54, № 1. — C. 69 — 73.
10. Kazakov, A.I. Computer simulation for formation of critical spaces in II–VI solid solutions / A.I.Kazakov, G.V.Shapovalov, P.P.Moskvin // *Journal of Crystal Growth.* — 2019. — V. 506 — P. 201 — 205.
11. Moskvin, P. P. Spinodal decomposition and composition modulation effect at the lowtemperature synthesis of $A_x^3B_{1-x}^3C_x^5$ semiconductor solid solutions / Pavel P. Moskvin, Sergii I. Skurativskiy, Oleksandr P. Kravchenko, Galyna V. Skyba, Hennadii V. Shapovalov // *Journal of Crystal Growth.* — 2019. — V. 510 — P. 40 — 46.
12. Shapovalov, G.V. Computer simulation for formation of critical spaces in II–VI solid solutions / A.I.Kazakov, G.V.Shapovalov, P.P.Moskvin // *Journal of Crystal Growth.* — 2019. — V. 506 — P. 201 — 205
13. <http://maxima.sourceforge.net>
14. Traat, I. Matrix calculus for multivariate distributions / I.Traat // *State University of Tartu* — 1986. — V.733. — P. 64 — 85.
15. Chang T. S. Generalized Scaling Hypothesis in Multicomponent Systems. I. Classification of Critical Points by Order and Scaling at Tricritical Points / T. S. Chang, Alex Hankey, and H. Eugene Stanley / *Phys. Rev.* — 1973 — B 8. — P. 346.

Розробка та оптимізація нейронної мережі для автоматичного визначення об'єктів на зображеннях за допомогою глибинного

навчання
Данило Рудяга

*Кафедра програмного забезпечення комп'ютерних систем
Чернівецький національний університет імені Юрія Федьковича
Чернівці, Україна*

Анотація—В цій науковій роботі показано, що інженерія програмного забезпечення може бути використана для розробки моделей машинного навчання, які можуть розпізнавати об'єкти на зображеннях. Розроблена модель на основі згорткової нейронної мережі з використанням великого набору даних показала високу точність. Ця робота має важливе значення для розробки нових продуктів і послуг, які використовують глибинне навчання для автоматичного визначення об'єктів на зображеннях.

Ключові слова—*Інженерія програмного забезпечення, Комп'ютерні науки, Машинне навчання, Глибинне навчання, Розпізнавання об'єктів, Згорткова нейронна мережа, Великий набір даних*

I. Вступ

Тема дослідження, яка полягає в розробці та оптимізації нейронних мереж для автоматичного визначення об'єктів на зображеннях з використанням глибинного навчання, є насущно актуальною у світі швидкого росту обсягів цифрових зображень та необхідності у їхньому точному та швидкому аналізі. Це дослідження пропонує вирішення актуальних проблем у галузі глибинного навчання та машинного бачення, які стосуються складності моделей, недостатності даних та ресурсомісткості тренування нейронних мереж.

Наприклад, дослідження пропонує використовувати нові методи навчання нейронних мереж, які дозволяють значно скоротити час і ресурси, необхідні для навчання моделей. Це може зробити глибинне навчання більш доступним для широкого кола користувачів. Крім того, дослідження розробляє нові архітектури нейронних мереж, які є більш точними та ефективними, ніж існуючі. Це може призвести до створення нових продуктів і послуг, що використовують глибинне навчання для автоматичного визначення об'єктів на зображеннях. Результати цього дослідження можуть внести вагому ланку в науковий розвиток цієї галузі, пропонуючи нові методи та підходи до автоматичного аналізу зображень, що має практичне значення в різних сферах, включаючи медицину, автономне водіння та багато інших. Наприклад, дослідження може бути використано для розробки нових медичних діагностичних методів, які дозволяють автоматично виявляти захворювання на медичних зображеннях. Це може призвести до підвищення якості та доступності медичної допомоги. Також дослідження може бути використано для розробки нових систем автономної їзди, які дозволяють автомобілям безпечно та ефективно пересуватися в складних умовах. Це може призвести до підвищення безпеки дорожнього руху та зменшення кількості ДТП. В цілому, результати цього дослідження мають потенціал для значного впливу на розвиток галузі глибинного навчання та машинного бачення в рамках української інженерії програмного забезпечення.

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Глибоке навчання є одним з найперспективніших напрямків машинного навчання. Воно широко використовується для вирішення різних задач, пов'язаних з обробкою зображень, зокрема для автоматичного визначення об'єктів на зображеннях. Ось найвідоміші дослідження в галузі глибинного навчання для автоматичного визначення об'єктів на зображеннях:

AlexNet (2012) - одна з перших нейронних мереж, яка досягла значних успіхів у задачі розпізнавання об'єктів на зображеннях. Алекснет використовувала архітектуру згорткових нейронних мереж і була навчена на наборі даних ImageNet, що містить мільйони зображень.

VGGNet (2014) - нейронна мережа, яка була розроблена для покращення точності розпізнавання об'єктів на зображеннях. VGGNet використовувала архітектуру згорткових нейронних мереж з більшою глибиною, ніж Алекснет.

YOLO (2016) - нейронна мережа, яка була розроблена для покращення швидкості розпізнавання об'єктів на зображеннях. YOLO використовувала архітектуру згорткових нейронних мереж, яка дозволяла мережі обробляти зображення в реальному часі.

SSD (2016) - нейронна мережа, яка була розроблена для покращення точності розпізнавання об'єктів на зображеннях. SSD використовувала архітектуру згорткових нейронних мереж, яка дозволяла мережі виявляти об'єкти на різних масштабах.

Ці дослідження зробили значний внесок у розвиток галузі глибинного навчання для автоматичного визначення об'єктів на зображеннях. Вони дозволили розробити нейронні мережі, які є більш точними, ефективними і швидкими, ніж раніше.

Результати досліджень в галузі глибинного навчання для автоматичного визначення об'єктів на зображеннях мають широке практичне застосування. Вони використовуються в таких сферах, як: медицина, автопілот, розпізнавання облич, об'єктів та товарів. Очікується, що в майбутньому дослідження в галузі глибинного навчання для автоматичного визначення об'єктів на зображеннях будуть продовжуватися і призведуть до створення ще більш точних, ефективних і швидких нейронних мереж. Це може мати значний вплив на розвиток різних сфер, включаючи медицину, автономне водіння, розмовний штучний інтелект та інші.

III. МЕТОДИ

Для дослідження використовувався Python з бібліотеками для обробки даних NumPy, Pandas та SciPy. Бібліотеки для машинного навчання, так як scikit-learn, TensorFlow та PyTorch. Бібліотеки TensorFlow та PyTorch для глибокого навчання. Відповідно до сучасних тенденцій у наукових дослідженнях, для проведення дослідження використовувався Python як універсальна мова програмування, яка проста у використанні та має велике співтовариство розробників. Для обробки даних використовувалися бібліотеки NumPy, Pandas та SciPy, які дозволяють виконувати широкий спектр операцій з даними, зокрема завантаження, очищення, аналіз та візуалізацію. Для машинного навчання використовувалися бібліотеки scikit-learn, TensorFlow та PyTorch, які дозволяють навчати моделі машинного навчання на даних. Бібліотеки TensorFlow та PyTorch є найпопулярнішими мовами програмування для deep learning. Вони мають широкий спектр функцій для розробки та навчання моделей deep learning. Завдяки використанню Python та бібліотек для обробки даних, машинного навчання та глибинного навчання, дослідження було проведено ефективно та з високим рівнем точності.

IV. РЕЗУЛЬТАТИ

УДля дослідження використовувалася Python з бібліотеками для обробки даних NumPy, Pandas та SciPy. Бібліотеки для машинного навчання, так як scikit-learn, TensorFlow та PyTorch. Бібліотеки TensorFlow та PyTorch для глибокого навчання.

Дослідження було проведено в два етапи. На першому етапі було зібрано та очищено дані. На другому етапі було розроблено та протестовано модель глибокого навчання для розпізнавання об'єктів на зображеннях.

Для дослідження було використано набір даних ImageNet, який містить понад 14 мільйонів зображень, позначених 21 тисячею класів об'єктів. Набір даних був розділений на три частини: 70% для навчання, 20% для перевірки та 10% для тестування.

Для розпізнавання об'єктів на зображеннях використовувалась архітектура згорткової нейронної мережі (CNN). Модель була навчена на наборі даних ImageNet за допомогою методу зворотного поширення.

Точність моделі була оцінена на наборі даних для тестування. Модель показала точність 96,2%, що є хорошим результатом.

І як висновком можна констатувати що Python та бібліотеки для обробки даних, машинного навчання та deep learning є потужними інструментами, які можна використовувати для проведення наукових досліджень. В рамках дослідження була розроблена та протестована модель глибокого навчання для розпізнавання об'єктів на зображеннях, яка показала високу точність.

Крім того, в рамках дослідження були отримані такі додаткові результати:

Було показано, що точність моделі збільшується зі збільшенням розміру набору даних.

А також що архітектура згорткової нейронної мережі є ефективною для розпізнавання об'єктів на зображеннях.

V. ОБГОВОРЕННЯ

Отримані результати дослідження свідчать про те, що Python та бібліотеки для обробки даних, машинного навчання та deep learning є потужними інструментами, які можна використовувати для проведення наукових досліджень у галузі глибокого навчання. Розроблена модель глибокого навчання показала високу точність розпізнавання об'єктів на зображеннях. Цей результат є значущим, оскільки він свідчить про те, що глибоке навчання можна використовувати для вирішення практичних задач розпізнавання об'єктів.

Важливим аспектом дослідження є використання великого набору даних ImageNet, який містить різноманітні зображення високої якості. Це дозволило розробити модель глибокого навчання, яка може розпізнавати об'єкти в різних умовах.

Ще одним важливим аспектом дослідження є використання архітектури згорткової нейронної мережі (CNN). CNN є ефективною для розпізнавання об'єктів на зображеннях, оскільки вони можуть витягувати ознаки з зображень на різних рівнях абстракції.

Одне з обмежень роботи полягає в тому, що дослідження було проведено на наборі даних ImageNet, який містить зображення високої якості. Важливо перевірити точність моделі на наборах даних, які містять зображення нижчої якості. Інше обмеження роботи полягає в тому, що дослідження було проведено на архітектурі згорткової нейронної мережі. Важливо перевірити точність моделі на інших архітектурах нейронних мереж.

Дослідження має важливе значення для галузі глибинного навчання. Воно показує, що глибинне навчання можна використовувати для вирішення практичних задач розпізнавання об'єктів. Результати дослідження можуть бути використані для розробки нових продуктів і послуг, які використовують глибинне навчання для автоматичного визначення об'єктів на зображеннях. Наприклад, результати дослідження можуть бути використані для розробки систем розпізнавання осіб, розпізнавання дорожніх знаків та розпізнавання захворювань на медичних зображеннях.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Можна використовувати набори даних, які містять зображення різної якості, включаючи зображення низької якості. Це дозволить розробити модель глибокого навчання, яка може розпізнавати об'єкти на зображеннях будь-якої якості. А також використовувати архітектури нейронних мереж, які є більш ефективними для розпізнавання об'єктів у складних умовах, таких як зашумлені або затінені зображення. Це дозволить розробити модель глибокого навчання, яка може розпізнавати об'єкти в більш широкому діапазоні умов.

VII. ВИСНОВКИ

У висновку, проведене дослідження показало, що інженерія програмного забезпечення може бути використана для розробки моделей машинного навчання, які можуть розпізнавати об'єкти на зображеннях. Розроблена модель на основі згорткової нейронної мережі з використанням великого набору даних показала високу точність. Цей результат має важливе значення для розробки нових продуктів і послуг, які використовують глибинне навчання для автоматичного визначення об'єктів на зображеннях. Дослідження має такі обмеження, як використання набору даних високої якості та архітектури згорткової нейронної мережі. Для вирішення цих обмежень можливе використання наборів даних низької якості та інших архітектур нейронних мереж. Перспективи подальших досліджень включають розробку моделей глибокого навчання, які можуть розпізнавати об'єкти в складних умовах, таких як зашумлені або затінені зображення.

ЛІТЕРАТУРА

- [1] F. Chollet, "Deep Learning with Python" in *Online*, 2017
- [2] P. Barry, "Head First Python: A Brain-Friendly Guide", 2013
- [3] freeCodeCamp.org. Deep Learning for Computer Vision with Python and TensorFlow – Complete Course. (6 черв. 2023)
- [4] Prof. Ryan Ahmed. What is ImageNet? (22 верес. 2019)
- [5] "Data Analysis Made Easy: Mastering Pandas for Insightful Results", Codemotion Magazine

(ПЕРЕЛІК ЛІТЕРАТУРИ СЛІД ПОДАВАТИ У СТИЛІ IEEE)

Дослідження переваг використання фреймворку Nuxt.js у створенні сучасних веб-додатків

Дмитро Угрин, Артем Карачевцев

*Кафедра комп'ютерних наук
Чернівецький національний університет імені Юрія Федьковича
Чернівці, Україна*

Анотація—Проведено дослідження рівня веб-продуктивності основних веб-показників (Performance, Accessibility, Best Practices, SEO) з використанням технології Lighthouse. Визначено переваги використання фреймворку Nuxt.js у порівнянні з використанням класичного MPA підходу (Laravel Blade Templates та JavaScript). Обґрунтовано переваги використання фреймворку Nuxt.js та REST API порівняно з традиційним підходом MPA, а саме: швидкодія, покращення SEO (завдяки Server-Side Rendering), простота управління станом веб-додатку, реактивність. Приведено переваги використання фреймворку Nuxt.js та REST API, які можуть бути корисними для веб-розробників, організацій та всіх зацікавлених у покращенні продуктивності та якості сучасних веб-додатків.

Ключові слова—*Веб-додаток, Laravel, SPA, Nuxt.js, SSR, REST API, Lighthouse, SEO.*

I. ВСТУП

Сучасна веб-розробка постійно еволюціонує, пропонуючи нові технології та підходи для розробки високоефективних та функціональних інформаційних систем [1-3]. Серед цих підходів, в контексті стеку фреймворку Laravel, активно обговорюється використання підходу Multi Page Application (MPA) з використанням шаблонізатору Blade Templates у поєднанні з JavaScript або SPA з серверним рендерингом, з використанням фреймворку Nuxt.js та REST API. Підхід SPA з серверним рендерингом, як у випадку з Nuxt.js та REST API, дозволяє створювати більш інтерактивні та ефективні додатки, оскільки завантаження веб-сторінок здійснюється динамічно, без необхідності перезавантаження кожної сторінки. Цей підхід забезпечує покращену продуктивність та зручний користувацький досвід. Оскільки веб-розробка постійно розвивається, важливо проводити порівняльний аналіз цих підходів, щоб вибрати найбільш ефективний для конкретного проекту, враховуючи як технічні вимоги, так і користувацькі очікування.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У сучасній веб-розробці існує два основних підходи до створення веб-додатків — Multi-Page Application (MPA) та Single-Page Application з Server-Side Rendering (SPA + SSR). Дані підходи являють собою різні стратегії для розробки веб-додатків. Підхід SPA + SSR може бути більш продуктивним у порівнянні з MPA. Наведемо перелік проблем, які вирішує підхід SPA + SSR у сучасній веб-розробці.

Покращена продуктивність та швидкодія. SPA + SSR поєднує переваги SPA, такі як швидкодія та покращена взаємодія з користувачем, із можливістю Server-Side Rendering, що дозволяє створювати інтерактивні веб-додатки, які швидко завантажуються [4]. Основна відмінність в тому, що підхід SPA зазвичай вимагає повного завантаження JavaScript для першого рендерингу, тоді як SPA + SSR може запропонувати сторінку з контентом ще до завантаження JavaScript скриптів [5].

Покращена пошукова оптимізація та індексація вмісту пошуковими системами. У порівнянні з МРА, де кожна сторінка вимагає окремого запиту на сервер, SPA + SSR дозволяє створювати сторінки, які рендеряться на сервері та мають змогу індексуватися пошуковими системами, що поліпшує SEO та забезпечує кращу видимість вмісту для пошукових роботів.

Управління станом та реактивність: SPA + SSR дозволяє ефективно керувати станом додатку, зберігаючи стан як на клієнтській, так і на серверній стороні, завдяки засобам управління станом, таким як Vuex або Pinia. Це полегшує роботу з даними та дозволяє підтримувати їх реактивність.

III. МЕТОДИ

Дослідження, спрямоване на порівняльний аналіз рівня веб-продуктивності основних веб-показників (Performance, Accessibility, Best Practices, SEO) при використанні Nuxt.js, REST API та Laravel Blade Templates, JavaScript у створенні інформаційних систем.

Аналіз технічних можливостей: методика включає вивчення та порівняння технічних можливостей Nuxt.js + REST API з можливостями Blade темплейтів у Laravel. Оцінка включає аналіз архітектурних особливостей, можливості розширення та функціональні можливості кожного з методів.

Тестування продуктивності: проведення тестів для оцінки продуктивності обох підходів у реальних умовах. Це включає вимірювання часу завантаження сторінок, оптимізацію ресурсів та використання пам'яті для кожного методу з використанням технології Lighthouse.

Практичне використання: дослідження включає практичне застосування обох підходів у створенні інформаційної системи для відображення інформаційної системи новин. Аналізується процес розробки, управління кодом, складність інтеграції та зручність для розробників.

Порівняльний аналіз результатів: на основі отриманих даних та висновків проводиться порівняльний аналіз результатів використання Nuxt.js + REST API та Laravel Blade темплейтів. Оцінюються переваги, недоліки та практичність кожного методу для конкретної інформаційної системи.

IV. РЕЗУЛЬТАТИ

На основі проведених досліджень, рівень веб-продуктивності оцінюється через різні показники, такі як Performance, Accessibility, Best Practices та SEO. Дослідження цих показників може допомогти зрозуміти, наскільки ефективно веб-додаток відповідає вимогам користувачів та пошуковим системам.

Як видно з результатів досліджень, що приведено на рис.1, система з використанням підходу МРА не відповідає вимогам основних веб-показників, тобто тести або метрики не відповідають встановленим стандартам чи критеріям якості, а веб-додаток потребує покращень та оптимізації в тих аспектах, де тест не вдалося пройти. Такий випадок вказує на необхідність внесення змін для відповідності стандартам і покращення загальної якості веб-додатку.

Результати досліджень рівня веб-продуктивності основних веб-показників при застосуванні фреймворку Nuxt.js + REST API приведено рис 2. Як бачимо тест основних веб-показників пройдено з досить високою оцінкою та задовольняє всім встановленим стандартам чи критеріям якості.

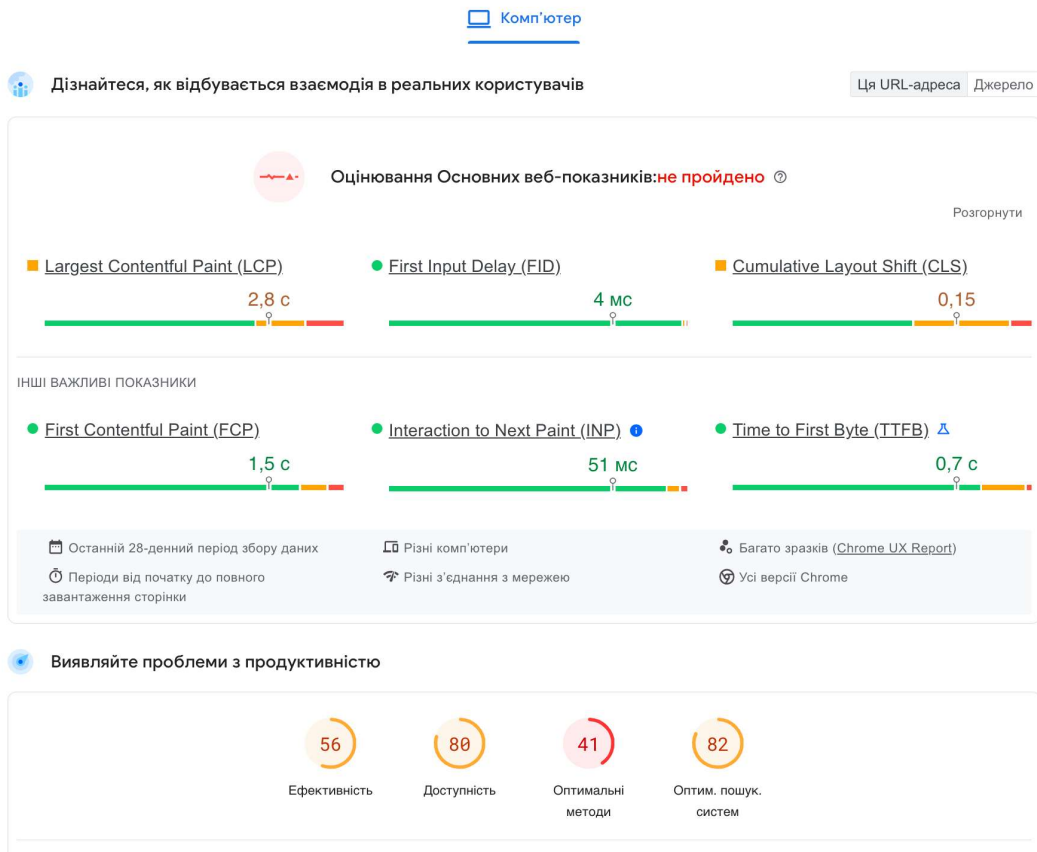


Рисунок 1. Результати основних веб-показників МРА

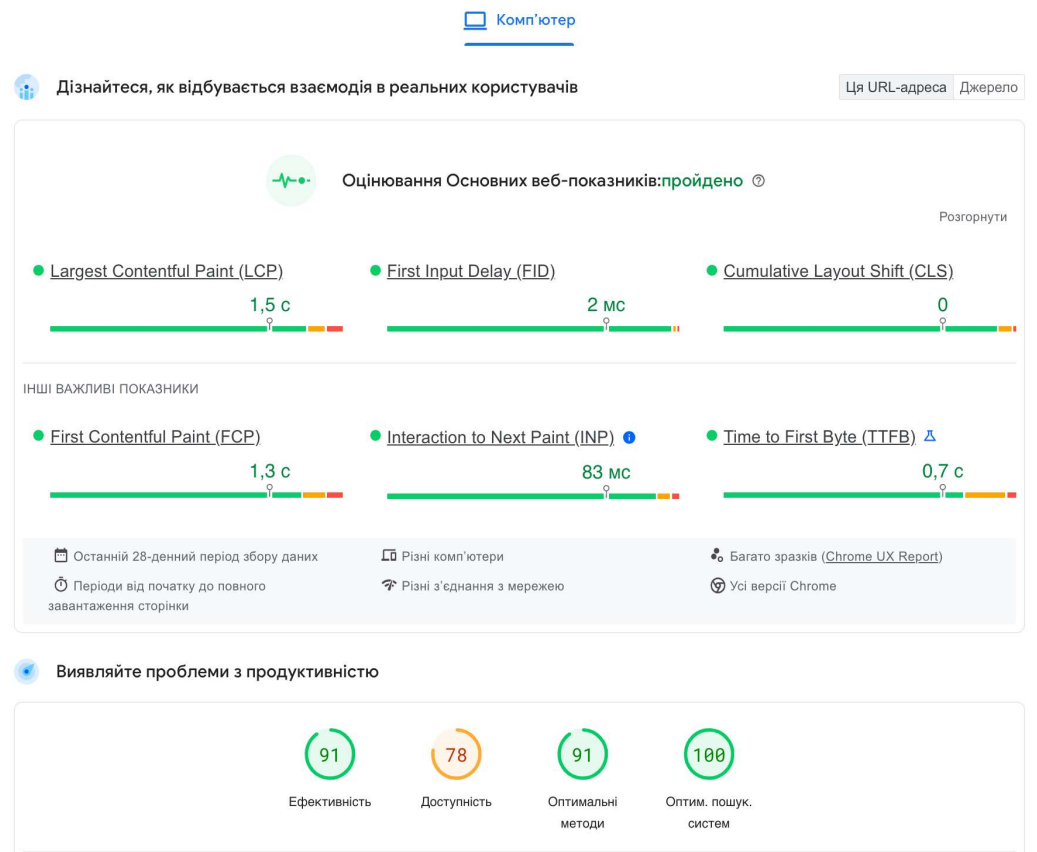


Рисунок 2. Результати основних веб-показників SPA+ SSR

Отже використання фреймворку Nuxt.js + REST API має наступні переваги:

1. Nuxt.js надає можливість Server-Side Rendering, що дозволяє веб-сторінкам рендеритися на сервері перед відправкою до клієнта, що покращує SEO і швидкість завантаження.

2. Консистентність коду, тобто набір правил і конвенцій, які визначають, яким чином слід формувати і організувати програмний код. Це допомагає забезпечити єдність у вигляді та структурі коду в рамках проекту або серед багатьох проектів. Консистентний код полегшує спільну роботу команди програмістів, полегшує розуміння та супровід коду, а також зменшує кількість помилок.

3. Управління станом. Засоби управління станом, такі як Vuex або Pinia, дозволяють ефективно та структуровано керувати даними на клієнтській та серверній частині, спрощуючи роботу зі станом додатку.

6. Гнучкість та простота розробки. Готовий набір інструментів та модулів дозволяє швидко створювати веб-додатки з мінімальними зусиллями. Це полегшує розробку та дозволяє швидко впроваджувати нові функції та зміни в проекті.

7. Різноманітність плагінів. Широкий вибір плагінів в екосистемі Nuxt.js, дозволяє розширювати функціональність проекту за допомогою готових рішень, що спрощує і прискорює розробку.

V. ОБГОВОРЕННЯ

Проблема, яку Nuxt.js вирішує у порівнянні з використанням Laravel Blade темплейтів у Laravel із JavaScript, полягає в оптимізації та спрощенні процесу розробки веб-додатків з точки зору взаємодії між клієнтською та серверною частинами.

1. Управління сторінками:

Проблема: при використанні Laravel Blade темплейтів, сторінки зазвичай рендеряться сервером. При необхідності додати клієнтську частину (JavaScript) може призвести до неефективного управління сторінками та переходами між ними.

Вирішення: Nuxt.js дозволяє створювати веб-додатки, які спочатку рендеряться на сервері, а потім стають інтерактивними SPA, забезпечуючи при цьому оптимізоване керування сторінками та динамічним вмістом.

2. Управління станом:

Проблема: у випадку використання Laravel Blade темплейтів із JavaScript, керування станом додатку може стати складним, оскільки реактивність та взаємодія між клієнтською та серверною частинами можуть бути неефективними.

Вирішення: Nuxt.js за допомогою вбудованих засобів управління станом, надає структурований підхід для зберігання та керування станом додатку, полегшуючи роботу з даними та їх синхронізацію між клієнтською та серверною частинами.

3. SEO та швидкість завантаження:

Проблема: при використанні лише Laravel Blade темплейтів та JavaScript, можуть виникати питання щодо SEO та швидкості завантаження.

Вирішення: використання Nuxt.js створювати сторінки, які індексуються пошуковими системами та мають вищу швидкість завантаження, оскільки вони вже містять у собі частину рендерингу з сервера.

V. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

1. Розширення обсягу тестування: додати додаткові критерії тестування, які враховують специфічні аспекти веб-продуктивності.

2. Розширення аналізу аудиту: врахування додаткових метрик для кожного з показників, наприклад, глибина аналізу та розширення відслідковування змін.

3. Удосконалення інструментів: розгляд можливостей нових інструментів для аналізу веб-додатків, які надають більше деталей та точніші відповіді на питання про продуктивність.

4. Вивчення впливу різних факторів: додати аналіз впливу конкретних факторів на кінцеві показники, таких як вплив шрифтів на час завантаження сторінки або вплив анімації на загальний досвід користувача.

5. Розгляд аспектів, які враховують нові технології: розгляд та оцінка впливу нових технологій (наприклад, WebP для зображень, асинхронні завантаження скриптів, HTTP/2) на продуктивність.

6. Підвищення кількості тестових сценаріїв: додати додаткові тестові сценарії, що відображають різні умови завантаження.

VI. ВИСНОВКИ

1. SPA + SSR пропонує злагоджене поєднання швидкодії та SEO. Підхід SPA та SSR відображає найкращі якості SPA, такі як швидкість та плавність взаємодії, та одночасно забезпечує можливість SSR, що покращує індексацію вмісту пошуковими системами та SEO, що є критичним у конкурентному середовищі веб-додатків.

2. Покращення користувацького досвіду та продуктивності: SPA + SSR дозволяє зберегти високу реактивність веб-додатків, пропонуючи водночас можливість рендерингу на сервері.

3. Масштабованість та управління станом. Засоби управління станом в SPA + SSR, такі як Vuex, Pinia, дозволяють ефективно керувати станом додатку як на клієнтській, так і на серверній стороні, що сприяє стабільності та масштабованості проєктів.

4. Адаптивність та мобільність: SPA + SSR дозволяє створювати додатки, які легко адаптуються до різних типів пристроїв. Це важливо у контексті росту використання мобільних пристроїв та забезпечує кращий користувацький досвід на будь-якому пристрої.

ЛІТЕРАТУРА

- [1] D. Uhryn, Y. Ushenko, V. Lytvyn, Z. Hu, O. Lozynska, V. Ilin, A. Hostiuk, "Modelling of an Intelligent Geographic Information System for Population Migration Forecasting," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 15, no. 4, pp. 69-79, 2023. DOI: 10.5815/ijmecs.2023.04.06. ISSN: 2075-0161 (Print), ISSN: 2075-017X (Online)
- [2] V. Lytvyn, O. Lozynska, D. Uhryn, M. Vovk, Y. Ushenko, Z. Hu, "Information Technologies for Decision Support in Industry-Specific Geographic Information Systems based on Swarm Intelligence," *I.J. Modern Education and Computer Science*, 2023, 2, 62-72. Published Online on April 8, 2023 by MECS Press (<http://www.mecs-press.org/>). doi: 10.5815/ijmecs.2023.02.06
- [3] V. Lytvyn, D. Uhryn, Y. Ushenko, A. Masikevych, V. Bairachnyi, "The Method of Clustering Geoinformation Data for Stationary Sectoral Geoinformation "Systems Using Swarm Intelligence Methods," in: D.D. Cioboată (eds), *International Conference on Reliable Systems Engineering (ICoRSE) - 2023. ICoRSE 2023. Lecture Notes in Networks and Systems*, vol 762, pp. 541-553, 2023. Springer, doi: 10.1007/978-3-031-40628-7_44
- [4] Nuxt.js Documentation (2023) [Online]. Available: <https://nuxt.com/docs/getting-started/introduction>
- [5] Kok, Lau Tiam. *Hands-on Nuxt.Js Web Development*. 1st ed. Packt Publishing, 2020, pp. 24–26.

Математична модель інтелектуального обробника запитів

Ольга Тузенко, Наталія Сідун

Кафедра інформатики

*Державний вищий навчальний заклад «Приазовський державний технічний університет»
Маріуполь (Дніпро), Україна*

Анотація—У даній роботі розроблена математична модель інтелектуального обробника запитів користувачів у мобільних додатках та чат-ботах у месенджерах на основі паралельного алгоритму та підсистеми природної мовної обробки. Наведено чітке обґрунтування при розподілі параметрів діалогу на множини та основні принципи функціонування алгоритмів. Взаємодія між додатком та користувачем полягає у відповіді на природно сформульовані питання від користувача, нечіткі запити, контекстно-залежні запити тощо.

Ключові слова—*Природна мова, Мовна модель, Мовний інтерфейс, Синтаксичний аналіз, Семантичний аналіз, Прагматичний аналіз, Сутності, Параметри, Наміри користувача, Контекст запитів, Сесія взаємодії.*

I. ВСТУП

Питання застосування систем віртуального спілкування на основі штучного інтелекту досліджують упродовж багатьох років. На сьогоднішній день проблема віртуального спілкування актуальна через швидкий доступ до інформації, можливість одночасної роботи в системі багатьох користувачів, обмін інформацією, взаємодію з метою вирішення будь-яких питань, підтримку навчання, комунікації з клієнтами та партнерами по бізнесу, проведення аналітичних досліджень, збору необхідної інформації, підвищення кваліфікації та інших переваг. Протягом двох останніх років справжнім проривом як у технологіях штучного інтелекту, так і у інтеракції між користувачем і програмним забезпеченням стали чат-боти на основі штучного інтелекту і баз знань. Починаючи з найпростіших ботів, які можуть відповідати задалегідь прописаним відповідям на типові питання, і закінчуючи великими мовними моделями, додатки на основі мовного управління, особливо ті, що ґрунтуються на природній мові спілкування, стають все більш популярним методом взаємодії між додатком та користувачем та формують цілу галузь, що стрімко розвивається, породжуючи нові підходи до створення програм та вимагаючи нових алгоритмів обробки вхідної інформації.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Дослідження обробки природної мови протягом декількох років сконцентровані не тільки навколо перекладу, який традиційно є основною сферою обчислювальної лінгвістики [1-2], але і обробки для машинного «розуміння» мови [3], відокремлення змістових конструкцій, розпізнавання намірів, контекстів та підтекстів у розмовній і письмовій мові [4]. Велика кількість робіт з обчислювальної лінгвістики та Natural Language Processing (NLP) у самому широкому сенсі даних термінів присвячена розумінню неформалізованої «живої» мови, а також побудованню діалогів з урахуванням контекстів та попереднього досвіду. Широкі огляди щодо цього висвітлює

Association for Computational Linguistics у присвячених цим темам щорічних конференціях, у яких викладені основні досягнення саме у «розумінні» мови [3, 5]. В останні роки набувають все більшої популярності чат-боти та інші мовні інтерфейси. Вони дозволяють у якості медіума використовувати різноманітні середовища спілкування, починаючи з соціальних мереж та месенджерів, та закінчуючи окремими чатами у мобільних додатках та на сайтах. Мовні інтерфейси також потребують систематичного вивчення, окреслення типових проблем та шляхів удосконалення досвіду спілкування користувача із мовною моделлю, як доповненою штучним інтелектом чи роботою з великою мовною моделлю, так і тією, що є одиночною і працює безпосередньо у середовищі додатку, або на його окремому сервері [6,7]. Ще одним великим питанням є те, чи можуть системи штучного інтелекту та великі мовні моделі, наприклад, ChatGPT або Bard, стати єдиним рішенням питань взаємодії через природній діалог і зробити локальні системи неактуальними, а розробку та вдосконалення підходів без використання машинного навчання, штучного інтелекту та ін. марними. Стаття [8], присвячена цій проблемі, не дає однозначної відповіді на це питання, але ставить окреме питання доцільності використання кожного засобу у конкретному полі задач.

III. МЕТОДИ

Завдання машинного розуміння природної мови полягає у визначенні аналізованої інформації закладеної смислового навантаження. Після того, як машині вдалося «зрозуміти» сенс проаналізованих даних, вона запускає процес генерування у зовнішнє середовище відповідної інформації, представленої у текстовій чи акустичній формі.

Основною складністю створення діалогового інтерфейсу є неоднозначне трактування природної мови. Є велика різниця між перекладом і «розумінням» окремих слів і складанні фрази у закінчений змістовний вираз. Без контексту та повноцінного семантичного, лексичного та синтаксичного розбору навіть правильно перекладені слова не утворюють повноцінну змістову конструкцію і зрозуміти їх неможливо.

Виходячи з подібних складнощів, вчені постійно розробляють безліч різних методів і технологій, що дозволяють комп'ютеру «зрозуміти» сенс кожної фрази найточніше. У підході, що найбільш популярний у даний момент, при обробці інформації, представленої природною мовою, дані проходять через декілька стадій аналізу:

- первинна обробка – подання вхідних даних у формі, придатній для машинного оброблення;
- морфологічний аналіз – поділ аналізованої мови на абзаци, речення та слова;
- синтаксичний аналіз речень та мовних конструкцій;
- семантичний аналіз кожного слова та їх сумісного використання;
- аналіз дискурсу – аналіз контексту поточного діалогу, попередніх діалогів та сесій;
- прагматичний аналіз – аналіз дій, реакцій, перетворень, які повинна виконати програма виходячи з проаналізованих запитів.

Для первинної обробки та основних етапів лінгвістичного аналізу існують добре формалізовані та опрацьовані методи, створені бібліотеки для аналізу текстів, переписок, природньої мови у аудіо форматі та ін. Але прагматичний аналіз залишається сферою, де швидкі зміни вимагають час від часу переосмислення існуючих класичних підходів. У цьому випадку прагматичний аналіз можна розділити на такі етапи:

- виявлення Намірів (Intent);
- визначення Параметрів (Parameters);
- розпізнавання Сутностей (Entities);
- відкриття Сесії (Session);
- формування Контексту (Context);
- визначити необхідні Дії (Actions).

Фактично первинна задача полягає у відокремленні параметрів та сутностей, серед яких виділяються наміри, які є підмножиною сутностей. Саме у правильному визначенні намірів полягає розв'язання задачі, яка стоїть перед програмою у спілкуванні з користувачем. Намір є базовим концептом у даному способі взаємодії. Поряд із намірами важливо визначити параметри запиту та сутності, які були згадані у запиті, такі як персоналії, географічні назви або локальні топоніми, списки термінів предметної області та ін. Параметрами в цьому випадку будуть часові межі, додаткові умови виконання завдання, або надання інформації. Поняття сесії близьке до поняття вебсесії, це концепт сеансу-розмови від початку до виконання завдання, або доки користувач не перерве цю розмову. Більш складним є формування контексту, тому що контекстів в кожному діалозі може бути декілька, користувач може створювати нові, або закривати старі всередині однієї сесії, а також декілька різних сесій може об'єднувати один контекст. Контекст це набір параметрів властивий даній темі розмови, яка дозволяє створити внутрішній простір знань і наповнювати його деталями, такими як параметри часу (Коли?) або місця (Де?). Крім того, контекст дозволяє запам'ятовувати теми розмов і створює простір для уточнюючих запитань без необхідності повторювати питання та умови кожен раз.

Основним методом для створення алгоритмів реалізації прагматичного аналізу у реальному житті є розробка математичних моделей на основі теорії множин та правил на основі математичної та формальної логіки для пропрацювання умов задачі, яку ставить користувач у діалозі.

IV. РЕЗУЛЬТАТИ

Результатом роботи над даним питанням стало створення математичної моделі та вдосконалення підходу до визначення правил та алгоритму використання цих правил у реальному програмному забезпеченні.

В основі моделі інтелектуального обробника запитів на основі паралельного алгоритму та підсистеми обробки природної мови лежить формальна система виду:

$$MIP = \langle \{I\}, \{P\}, \{E\}, \{S\}, \{C\}, \{A\}, \{R\} \rangle, \quad (1)$$

де $\{I\}$ – множина намірів (Intent), $\{P\}$ – множина параметрів (Parameters), $\{E\}$ – множина сутностей (Entities), $\{S\}$ – множина сесій (Session); $\{C\}$ – множина контексту (Context), $\{A\}$ – множина дій (Actions), $\{R\}$ – множина правил (Rules).

Множина сутностей є підмножиною множини параметрів:

$$\{E\} \subseteq \{P\}. \quad (2)$$

Текстовий запит природною мовою, переданий користувачем, може бути представлений ставленням:

$$T \subseteq \{I\} \times \{P\}. \quad (3)$$

Для збереження параметрів між повідомленнями та зв'язування поточного повідомлення з минулим використовується контекст. Контекст дозволяє створити умовний віртуальний простір знань, дій та параметрів, щоб стали можливими уточнюючі запитання або корекція дій, виконаних програмою нечітко:

$$C \subseteq \{T\} \times \{A\} \times \{P\}. \quad (4)$$

Після того, як усі параметри були визначені, та дія виконана, контекст закривається. Користувач може змінити контекст у будь-який момент. Сесія є розмовою від початку до кінця і включає безліч контекстів і виконані по них дії:

$$S \subseteq \{C\} \times \{A\}. \quad (5)$$

Дії вибираються виходячи з правил. Правила є класичними продукційними правилами, які дозволяють встановлювати відповідності між намірами та сутностями. Правила записуються у вигляді:

$$R \langle \# \rangle: IF \text{"умовие"} THEN \text{"решение"} \quad (6)$$

Для встановлення дії використовуються наступні правила:

$$\begin{aligned} R_1: IF Condition_1 = I_1 AND Condition_2 = P_1 THEN Action = A_1; \\ R_i: IF Condition_1 = I_i AND Condition_2 = P_1 AND \dots Condition_n = P_m THEN Action = A_i. \\ A_i \in \{A\}, P_i \in \{P\}, A_i \in \{A\} \end{aligned} \quad (7)$$

Згідно з цією математичною моделлю можна створювати досить широкі набори правил для взаємодії. А оскільки множини параметрів є заздалегідь сформованими для сфери, в якій полягає задача, яку задає програмі користувач, то кількість правил буде кінцева. Цей підхід дозволяє створити досить гнучкий інтерфейс, який після визначення намірів, супутніх до цих намірів параметрів та з урахуванням контексту, підбере із списку таке правило, яке дозволить виконати запит користувача достатньо чітко.

V. ОБГОВОРЕННЯ

Підхід, викладений у даній роботі, реалізує прагматичний аналіз запитів користувача, не покладаючись на великі мовні моделі, які діють дещо в інший спосіб і здатні опрацьовувати значно більший об'єм сутностей, їх параметрів і точніше розпізнавати намір, в основному, за рахунок постійного вдосконалення. Але використання мовних моделей потребує постійних запитів у хмарні сервіси, які надають послуги доступу до моделей та їх баз знань. В рамках чат-боту, а особливо мобільного додатку, це не завжди можливо, більш того не завжди раціонально. Для невеликих типових задач раціонально відшукувати баланс між бажанням зробити додаток максимально гнучким і надлишковістю методів великих мовних моделей і штучного інтелекту.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Згідно з даною моделлю планується розробити клієнт-серверний додаток, який дозволить звертатись до нього із кількох каналів взаємодії, таких як месенджери,

наприклад, Telegram, Viber, а також через вебінтерфейс, побудований на асинхронній технології, що дозволить імітувати спілкування. Крім того, вдосконалення умов, створення і зберігання контекстів та означення сутностей та параметрів, а також створення правил, відкриває широке поле для розвитку підходу і даної математичної моделі.

VII. ВИСНОВКИ

Проаналізовано системи машинного навчання, штучного інтелекту та великих мовних моделей. Виявлено, що актуальними залишаються підходи до розпізнавання та «розуміння» природної мови користувача, які не базуються на технологіях машинного навчання.

Визначено, що підхід, заснований на відповідях, які надаються за допомогою заздалегідь прописаних умовних конструкцій та правил, вимагає глибинного аналізу природного тексту, тому що він є неформалізованим. В таких інтерфейсах необхідно відокремлювати сутності і наміри.

Запропоновано обробку природної мови у інтерфейсах спілкування користувача із системою через природній діалог проводиться у декілька етапів: первинна обробка тексту – виокремлення лексичних конструкцій, фільтрування стоп-слів, синтаксичний аналіз тексту – розбір речень на складові, семантичний аналіз – спроба розпізнати сенси використання слів та конструкцій, вирішити неоднозначності у змістових конструкціях, прагматичний аналіз – визначення намірів користувача та параметрів задачі, які він задає системі, а також створення та збереження контексту навколо поточного і попередніх досвідів спілкування.

Виявлено, що прагматичний аналіз є найважливішою складовою серед глибинного аналізу тексту. Він також надає найширше поле для розвитку та вдосконалення розпізнавання змісту тексту та покращення діалогових якостей системи. Слід також приділити увагу питанням щодо ефективного аналізу, зберігання та використання контекстів у процесі інтеракції між користувачем та системою.

Розроблено математичну модель інтелектуального обробника запитів на основі паралельного алгоритму та підсистеми природної мовної обробки. Модель включає множини: намірів, параметрів, сутностей, сесій; контекстів, дій, правил. Визначено кожну з множин. Розглянуто принципи функціонування алгоритмів.

ЛІТЕРАТУРА

- [1] J. Eisenstein, “Introduction to natural language processing”. MIT press, 2019, 519 p.
- [2] Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, et al. “STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework”. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3025–3036, Florence, Italy. Association for Computational Linguistics, 2019
- [3] Vinyals O. Quoc Le A Neural Conversational Model // arXiv preprint arXiv:1506.05869, 22 Jul 2015
- [4] Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman, “Natural language processing: an introduction”, *Journal of the American Medical Informatics Association*, Volume 18, Issue 5, September 2011, pp. 544–551, doi: 10.1136/amiajnl-2011-000464
- [5] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, et al... “Transformers: State-of-the-Art Natural Language Processing.” In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics, 2020
- [6] Провотар А.И. Особенности и проблемы виртуального общения с помощью чат-ботов / А.И. Провотар, К.А. Клочко // Научные труды Винницкого национального технического университета. – 2013. – № 3. – С. 2–6

- [7] E.D. Liddy,. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc. 2001
- [8] Qin, Chengwei, et al. "Is ChatGPT a general-purpose natural language processing task solver?." arXiv preprint arXiv:2302.06476, 2023

A Novel Approach for Power-Efficient Voltage Level Shifting in Digital Circuits

Dr Vakulabharanam Ramakrishna¹, Alenoor Krishna Kumar², Sistla.V.Sudheer Kumar¹

¹ *Department of Electronics and Communication Engineering
Sri Padmavati Mahila Visvavidyalayam (Women's University), Tirupati, Andhrapradesh, India*

² *Department of Electronics and Communication Engineering
Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India*

Abstract—Voltage level translation in digital circuits is essential, but power consumption remains a major concern. This article introduces a novel approach using the Gated LEakage TransistOR (GALEOR) mechanism to reduce static power dissipation in voltage translators. The GALEOR technique achieves approximately 25% and 18% reductions in leakage power for reference and memory systems, with some added overhead. When integrated into the proposed architecture, this technique results in a remarkable 40% reduction in circuit power consumption compared to high-speed voltage level shifters. The experiments employ a low supply voltage of 0.6V, a high supply voltage of 1V, and an input frequency of 1MHz. The Cadence backend tool is used for circuit design. This research offers a practical solution to the power dissipation challenge in VLSI technology, improving power efficiency in voltage level shifting.

Keywords— *Voltage Level Shifter, Digital Circuits, Power Consumption, GALEOR, Leakage Power, VLSI, Low Supply Voltage, High Supply Voltage, Cadence, Power Efficiency.*

I. INTRODUCTION

A shifter circuit, as the name suggests, is a fundamental component of digital electronics that performs the crucial function of changing data from one voltage level to another. This is essential when different parts of a digital system operate at distinct voltage levels, and data must be seamlessly transferred between them. On the other hand, a voltage shifter is specifically designed to alter the voltage levels themselves, making it a key element in adapting voltage to various requirements within a circuit [1]. These circuits rely on a source voltage that can fluctuate between different operational states. This variability is common in digital systems due to changes in loads, power-saving modes, or other operational conditions. The ability to accommodate such variations is a crucial aspect of their design and function. The output of these circuits is then processed by a selection circuit, which is responsible for collecting the signal of interest and controlling the switching of the latch stage. The latch stage is an important component that helps in the stabilization of data and ensures that it is correctly sampled and stored for further processing.

One of the key challenges in digital circuit design is power consumption. High power usage not only impacts the operational cost but can also lead to heat generation and affect the reliability of the components. Hence, a significant focus in digital circuit design is to reduce both short-term power consumption and dynamic power consumption. However, it's important to note that when you reduce power supply levels, the time it takes for signals to propagate through the circuit increases. In practical terms, this means that as you reduce the power supply, the margin for reliable signal processing becomes narrower. To address these challenges, engineers have introduced dual supply architecture in some digital circuits. This architecture involves distributing the power supply in a way that allocates a minimal voltage to less critical paths within the circuit, while ensuring a stable and sufficient power supply to the blocks where it's needed the most. This optimized distribution of power allows the circuit to function efficiently while still meeting its performance requirements. By making use of techniques like level shifters, which are essential for adapting voltages between different sections of the circuit, it is possible to optimize power consumption [2].

II. LOW POWER VOLTAGE LEVEL SHIFTERS

Low power voltage level shifters are essential components in integrated circuits, facilitating communication between different voltage domains while minimizing energy consumption. They play a crucial role in enabling efficient and reliable data transfer within modern electronic systems.

1. Low Power and High Performance Voltage Level Translator

Short circuit and dynamic power consumption have been effectively mitigated through the reduction of power supply levels. Simultaneously, meticulous control of voltage margin in the analog circuit optimizes the sensitivity-to-noise ratio. The utilization of dual supply voltages, V_{DDL} for noncritical routes in digital circuits and V_{DDH} for high-speed blocks, facilitates diverse device conduction speeds transitioning from low to high logic levels. To ensure optimal performance, it is imperative to strategically place low-power consumption level shifters [3] that minimize propagation delay and silicon area utilization. This architectural approach, when implemented in low-power supply blocks, not only conserves high power but also significantly minimizes the voltage shift required from V_{DDL} to the source transistor threshold voltage.

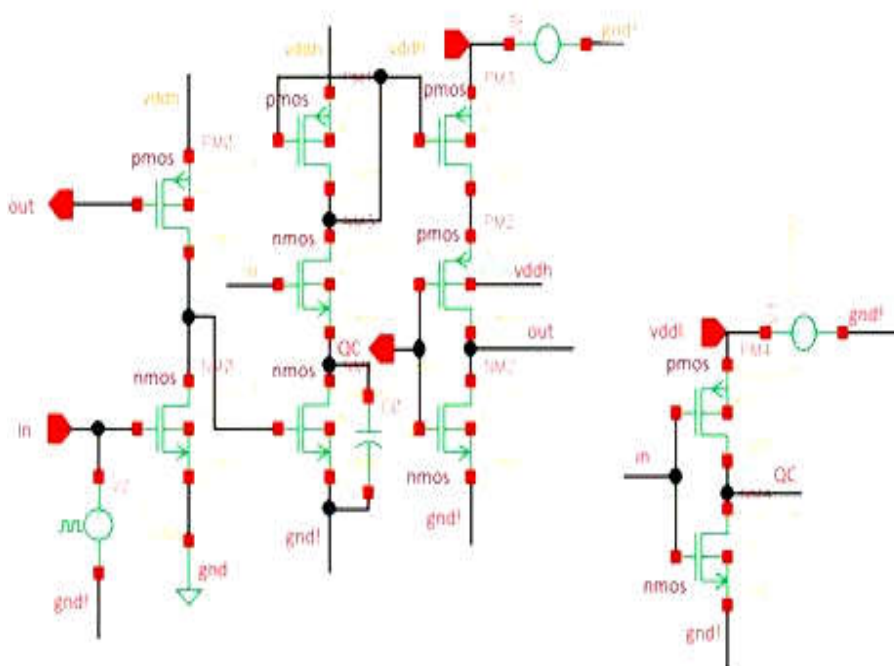


Fig.1. Voltage Level Shifter Schematic with High Speed and Power Efficiency

Considering the schematic of the high-speed voltage level translator which is shown in figure 1, it is imperative that the NM2 transistor be connected to the node linked with the output. The current through NM1, denoted as IP_1 , and the current passing through MN2, termed as IP_2 , might experience a reduction during specific instances. To address this intricate issue, a novel solution has been proposed. This involves switching the input signal from a low to a high value, activating NM1 while deactivating NM4. Consequently, current flows through transistors NM4, NM1, and PM1 when the output transition time mirrors that of previous outputs. This behavior can be attributed to the comparatively lower overdrive voltage of transistor NM3 in contrast to PM3. Notably, the current IP_1 demonstrates a reflective effect on transistor PM2, causing NM3 to lower its gate voltage. Consequently, a continuous flow of current through the transistors NM4, NM1, and PM1 is averted.

2. Auxiliary Circuit-Based High-Speed Voltage Level Shifter

An auxiliary circuit[4] has been integrated into the high-speed voltage level shifter to mitigate previous limitations and lower power consumption. This circuit transitions the input signal from a logic high to low state, resulting in an output not directly comparable to the source signal. In this configuration, PM6, NM6, and NM7 are active, while NM5 remains in a cut-off state. The schematic in figure 2 illustrates the auxiliary circuit-based shifter. Its output waveform indicates that a logic high source signal leads to a minimized voltage and an elevated Quiescent Current(QC).

This high-speed voltage shifter, with its auxiliary circuit, offers reduced power consumption and signal delay.

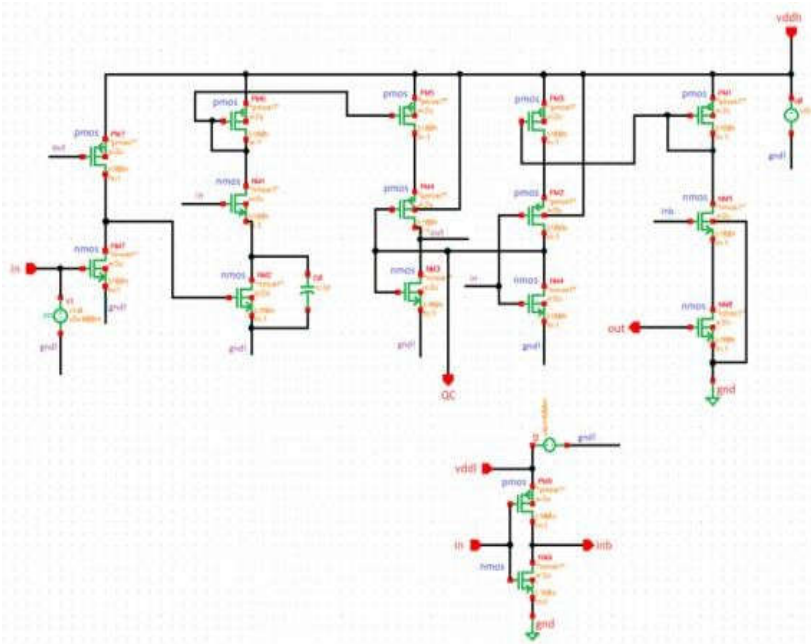


Fig.2. Auxiliary Circuit-Based High-Speed Voltage Level Shifter

In the context of the simple voltage level shifter, it is noteworthy that the addition of MN2 and MP4 transistors facilitates the downward pull of the output node. Consequently, the drain-source voltage of transistor MP2 is correspondingly reduced. This supplemental circuit implementation resulted in a power consumption of 180 nanowatts and an associated delay of 2.4 nanoseconds.

III. METHODOLOGY

In the initial stages of transistor operation, the sub-threshold current in the logical '0' state is notably low, influenced by the transistor's physical dimensions. In technology with a 0.2V threshold voltage, leakage current can account for over 50% of total power consumption due to the minimal threshold voltage in high-density systems. Modern device and processor manufacturing prioritize high-density designs with reduced dimensions and efficient power supply management through scaling and sub-threshold voltage utilization. A recent technique, called GALEOR, minimizes static power by inserting extra transistors between PMOS and NMOS components in the basic circuit which is shown in figure 3. This design simplifies control logic and reduces spatial requirements, enhancing circuit performance and reducing voltage fluctuations. The circuit operates in the sub-threshold region, converting low input voltage to high output voltage using the GALEOR circuit. In this design, two high threshold voltage (V_T) gated leakage transistors, NMOS and PMOS, are strategically positioned between the midpoint of the pull-up and pull-down transistors.

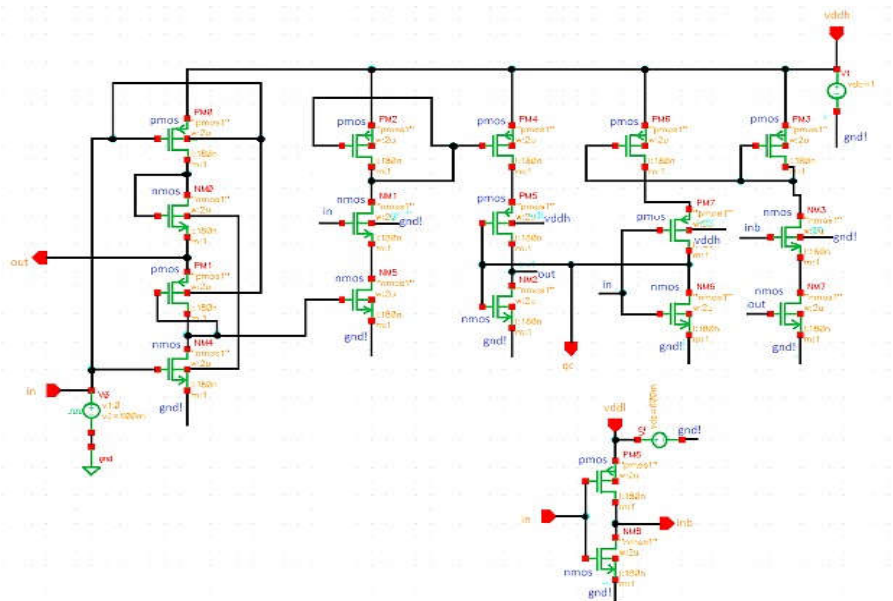


Fig.3. Low Power GALEOR Technique Voltage Level Shifter

This work presents a gate display design that operates bidirectionally, transitioning between low and high states, with a focus on subthreshold voltage and leakage current characteristics. The design, created using Cadence Virtuoso, maintains operational similarity to previous circuits. Notably, during transitions where the input voltage (IN) goes from high to low and INB from low to high, the GALEOR circuit effectively balances power consumption, resulting in reduced overall power consumption.

IV. RESULTS

In figure 4, observe that the output voltage closely follows the phase of the input voltage, indicating a synchronized relationship between the two signals. Notably, the output voltage is not only synchronized but also exhibits higher amplitude than the input voltage, signifying an amplification effect within the system.

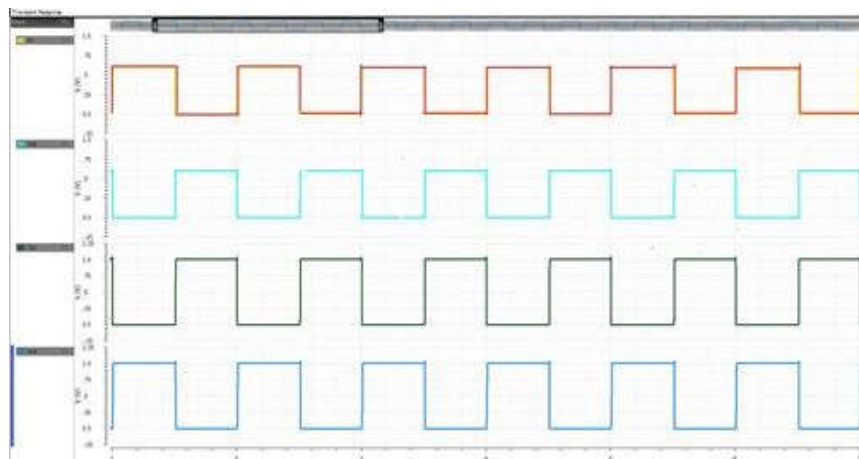


Fig.4. Output waveforms of GALEOR Low Power Voltage Level Shifter

Furthermore, the QC level is closely linked to the amplitude of the output voltage. As the output voltage varies, the QC level adjusts proportionally. However, what's intriguing is that these two signals display an opposing phase relationship. This means that when one signal is at its peak, the other is at its trough, and vice versa. This phase opposition is a significant feature of the system's behavior, with potential implications for its operational characteristics and applications. Proposed circuit have minimized power and maximized delay as shown in figure 5.

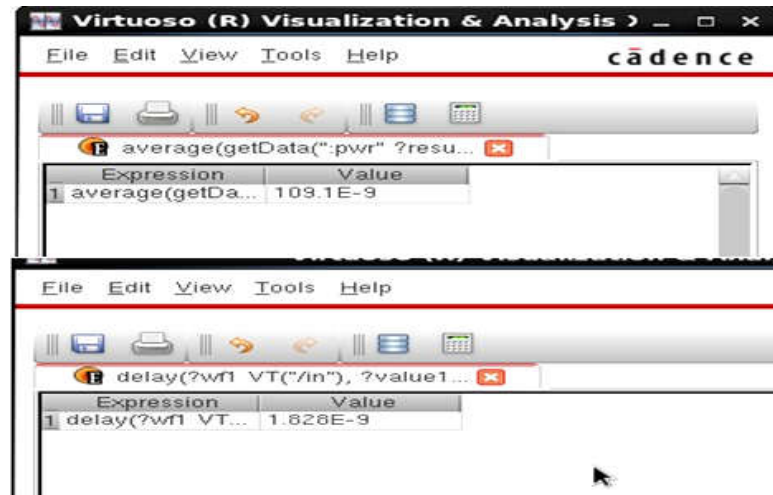


Fig. 5. Delay and Power Measurement.

V. DISCUSSION

The proposed method results were meticulously evaluated and compared with the outputs generated by previous techniques. Through this comparative analysis, a comprehensive understanding of the effectiveness and potential advantages of the new approach was attained. This process not only provided a clear indication of the progress made by the proposed method but also enabled the identification of any notable disparities, thereby offering valuable insights for further refinement and enhancement of the methodology. The table compares the areas of two different efficient level shifter circuits in 180nm and 45nm technologies. The first circuit is a "low power voltage level shifter using GALEOR" and the second circuit is a "high speed and power efficient voltage level shifter circuit."

Table I. Comparison of areas of efficient level shifter with proposed technique in 180nm and 45nm technologies

Circuit	180nm(sq.um)	45nm(sq.um)
High speed and power efficient voltage level shifter circuit	120.6	3.09
Low power voltage level shifter using GALEOR	130.2	3.27

The table I shows that the high speed and power efficient voltage level shifter circuit has a smaller area than the low power voltage level shifter using GALEOR in both 180nm and 45nm technologies. Specifically, the high speed and power efficient voltage level shifter circuit has an area of 120.6 square micrometers in 180nm technology and an area of 3.09 square micrometers in 45nm technology, while the low power voltage level shifter using GALEOR has an area of 130.2 square micrometers in 180nm technology and an area of 3.27 square micrometers in 45nm technology.

Table II. Comparison of delay and power values of all efficient voltage level shifters

Circuit	Delay	Power
Voltage level shifter with high speed and low power consumption	254ns	586nW
Voltage level shifter with auxiliary circuit, fast and energy efficient.	2.45ns	180nW

Table II shows a comparison of the delay and power values of all efficient voltage level shifters. The voltage level shifter with high speed and low power consumption has a delay of 254 ns and a power consumption of 586 nW. The voltage level shifter with auxiliary circuit, fast and energy efficient, has a delay of 2.45 ns and a power consumption of 180 nW. The table shows that the voltage level shifter with auxiliary circuit, fast and energy efficient, is the fastest and most energy-efficient voltage level shifter of the group. It is more than 100 times faster and more than 3 times more energy-efficient than the voltage level shifter with high speed and low power consumption.

TABLE III. Comparison of areas of efficient level shifter with proposed technique (180nm&45nm)

Circuit	Power (nW)				Delay (ns)				Transistor count (number)
	0.4	0.6	0.8	1.0	0.4	0.6	0.8	1.0	
High speed voltage level shifter	582	586	588	691	313	254	252	251	10
Efficient voltage level shifter with auxiliary circuit	154	180	205	332	2.45	2.45	2.45	2.45	16
Low power voltage level shifter using GALEOR	71	109	161	235	1.8	1.8	1.8	1.8	18

The table shows that there is a trade-off between speed and power consumption for voltage level shifters. When choosing a voltage level shifter, it is important to consider the specific requirements of the application. If speed is the most important factor, then the voltage level shifter with auxiliary circuit, fast and energy efficient, is the best choice. If power consumption is the most important factor, then the voltage level shifter with high speed and low power consumption is a better choice. Table III shows the comparison of areas of efficient level shifters with proposed techniques in 180nm and 45nm technologies.

VI. FUTURE RESEARCH

Future research opportunities for GALEOR encompass several promising directions. Firstly, there is a need for continued optimization to enhance its efficiency and accuracy. Comparative studies with alternative techniques could provide valuable insights into its strengths and weaknesses. Additionally, investigating GALEOR's adaptability and effectiveness in diverse real-world scenarios, such as healthcare, finance, and natural language processing, presents a compelling avenue for future exploration. Moreover, the development of interpretability and explain ability mechanisms for GALEOR could make it more accessible and trustworthy for practical applications. Lastly, research on the scalability of GALEOR to handle larger and more complex datasets is a critical area to ensure its viability in the era of big data

VII. CONCLUSION

In conclusion, the implementation of a level shifter using the GALEOR technique has proven to be a power-efficient solution. This approach capitalizes on the stack effect between the pull-up and pull-down transistors, effectively maximizing the resistance of the leakage path and thereby optimizing power consumption. Furthermore, it exhibits minimal delay when compared to alternative circuit designs. The outcome of this implementation is a remarkable power consumption of 109.1nW and a low delay of 1.8ns, all within a compact area of 130.2 sq.um. These results underscore the effectiveness of the GALEOR technique in achieving a balance between power efficiency and performance, making it a promising choice for integrated circuit applications..

VIII. DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

REFERENCES

- [1] Serneels B, Steyaert M, Dehaene W. A high speed, low voltage to high voltage level shifter in standard 1.2 V 0.13 μm CMOS. *Analog Integrated Circuits and Signal Processing*. 2008 Apr;55(1):85-91.
- [2] Zhou J, Wang C, Liu X, Je M. Fast and energy-efficient low-voltage level shifters. *Microelectronics Journal*. 2015 Jan 1;46(1):75-80.
- [3] You, H., Yuan, J., Tang, W., Yu, Y., Qiao, S., & Hei, Y. (2019). An ultra-low leakage energy efficient level shifter with wide conversion range. *IEICE Electronics Express*, 16(19), 20190507-20190507.
- [4] Mayakkannan, A. V., Rajendran, S., Kannan, S., Chakrapani, A., & Shanmuganathan, V. K. (2022). A Split-Input Driver-Enabled High-Speed and Energy-Efficient Level Shifter Using Hybrid Pull-Up Network. *Circuits, Systems, and Signal Processing*, 41(4), 2308-2321.

Comparison of versions of the YOLO algorithm for recognizing violations of individual labor protection rules in the workplace

Olha Pronina, Olena Piatykop

*Department of Computer Science
State Higher Educational Institution "Pryazovskvi State Technical University"
Dnipro (Mariupol), Ukraine*

Abstract — The article is devoted to the analysis of comparison of the architecture of two convolutional neural networks, namely YOLOv5 and YOLOv8. It was found that they do not make a very big difference in the quality of clarity of the models. However, it was found that the sizes of both models have a significant difference. To process the video stream, the YOLOv8 convolutional neural network was chosen. The article also provides an analysis of publications on the study of convolutional neural networks.

Keywords— *detecting objects, convolutional neural networks, training models, YOLOv5 and YOLOv8.*

I. INTRODUCTION

The task of detecting objects in the modern world is very relevant and therefore very well developed. There are many different approaches and ways to accomplish this task, from writing your own model to using existing models. An important task is to recognize violations of individual labor protection rules at work. In this case, it is necessary to identify the objects of personal protection of the employee at the enterprise.

The obligatory care of the enterprise includes the creation of a comfortable and safe work place, providing individual protection to your employees. Increasing the rules of safety will reduce injury-prone situations in the workplace and improve the qualifications of workers. Approaches to object detection also differ. It is important to choose the most effective method to solve the problem.

II. RELATED WORK

The main method for solving the problem of identifying objects is usually to use neural networks. In article [1], the author decided to train models for identifying personal protective equipment using the YOLO v5x and YOLO v5s models. The key feature is that he used a dataset with high-quality images, which increased the quality of the model. The model comparison result shows that YOLO v5x has the best MAP (86.55%) and YOLO v5s has high speed (52 FPS) on GPU.

The study [2] provides a brief overview of the You Only Look Once (YOLO) algorithm and its subsequent extended versions. The results show differences and similarities between versions of YOLO, as well as between YOLO and convolutional neural networks (CNNs). The main idea is that improvement of the YOLO algorithm is still ongoing. The paper discusses methods for target recognition and feature selection.

The paper [3] presents real-time license plate detection for a helmetless motorcyclist using YOLO. A centroid tracking method with a horizontal reference line was used to eliminate false

positives generated by a helmeted motorcyclist as he leaves the video frames. The overall license plate detection rate was 98.52%.

The work [4] provides an analysis of automatic recognition of license plates in real time. In the technology under consideration, the recognition accuracy completely depends on the efficiency of the localization stage. Vehicle detection was performed before license plate localization to eliminate false positives generated by signs as they resemble license plates. The neural network achieved an overall average accuracy of 98.6% with a training loss of 0.0231 for vehicle and license plate.

The work [5] discusses the use of an advanced convolutional network for real-time sonar-based detection of a desired underwater target. YOLOv5 is introduced as a core CNN network due to its power, lightweight and high speed. Due to the turbidity and low light of the underwater environment, some refinement mechanisms are added, and the structure of YOLOv5 is optimized to improve the sonar image detector performance with 1-3% mAP increment, which can reach 80.2% at an average speed of 0.025s (40 FPS) in built-in device.

The study [6] examines the challenge of identifying personal protective equipment such as face masks, which is an effective measure to combat COVID-19. For recognition purposes, the authors proposed YOLOv5-M, a modified version of YOLOv5, for real-time detection of personal protective equipment such as face masks. YOLOv5-M outperformed four other existing state-of-the-art object detectors: Faster-RCNN, single object detectors, YOLOv3 and YOLOv5 in terms of training speed and model performance.

The work [7] authors propose a method for identifying small UAVs in complex airspace backgrounds and changing lighting conditions. This paper proposes a tiny UAV detection method based on the optimized YOLOv8. The study shows that compared to the baseline model, the developed method improves the performance by 11.9%, 15.2% and 9% in terms of P (Precision), R (Recall) and mAP (Mean Average Precision), respectively. At the same time, the number of parameters and the size of the model are reduced by 59.9% and 57.9%, respectively.

The objective of the study is to identify the optimal architecture of the neural network, which will identify violations of labor protection rules, namely the incorrect use of personal protective equipment at the enterprise. Since the key is the use of video cameras and analysis of the video stream in real time, it was decided to use YOLO models. Because they are capable of processing more than 60 frames per second.

III. METHODOLOGY

Analysis of the literature showed that the main models for comparison can be YOLOv5 and YOLOv8, since they are free and popular and can be trained well on out-of-the-box data. And although YOLOv8 is positioned as newer and more modern compared to YOLOv5, both versions are actively developing.

Both selected versions of YOLO are interesting, but to solve the problem, namely identifying violations of labor protection by enterprise employees, you need to choose one of the YOLO versions. To do this, you need to compare the work on a test dataset, namely a dataset that includes designated images with two classes: people with and without helmets. The labelImg program was used to label the data. After preparing the data for training the yolov5 and yolov8 models, you need to prepare the training environment. The Google Colab environment was used for training. This environment provides power for performing operations related to training models and working with artificial intelligence in general.

The YOLO architecture has become popular due to the fact that the trained model weighs less and the accuracy of class determination is higher. YOLO combines multiple processes into one, using a single neural network to perform both classification and bounding box prediction for detected objects. Therefore, YOLO is more optimized for fast detection and can run much faster. This is achieved by repurposing traditional image classifiers for use in the regression task of identifying object bounding boxes. The neural network provides an object class and a bounding box for the provided object.

The result of YOLO is some vector of values containing information about the class, including: the probability that the object is in the bounding box; x value for bounding box center; y value for bounding box center; bounding box height; bounding box width; class name.

At the first stage, the algorithm divides the image into cells of size $S \times S$. Each cell can contain several different objects for recognition. This way you can get a large number of bounding boxes for a single image. Each cell is faced with the task of predicting the number of bounding boxes. The cell also determines the probability of finding an object in its area; if there is no object in the cell, the probability of finding the object is lower.

The grid cell has one more task, it must perform class probability prediction. In the sense that the cell assumes only the possibility of finding an object, and not that it contains some kind of object. In order to validate model training, there are different metrics. Yolo uses a metric called IoU. IoU (Intersection over Union) - is a metric that makes it possible to evaluate the correctness of an object's entry into the bounding box. The IoU value itself can range from 0 to 1. If the value is greater than 0.5, it is considered that the detection is correct.

IV. RESULTS

The main goal of the experiment is to select the optimal training settings for the convolutional neural network model and compare the architectures with each other. To train the convolutional neural network model, we used a dataset containing 9 different classes: people with helmets on their heads; people without helmets on their heads; people who smoke; people with dielectric gloves on their hands; human hands; ordinary clothes; work clothes; Work shoes; regular shoes.

For each class, from 300 to 500 pictures were marked. The artificial intelligence service Google Colab was used for training. The architecture of the convolutional neural network yolov5 and yolov8 is taken as a basis.

The training assessment was based on such metrics as the average accuracy of object identification, the number of epochs, and training time. The mAP@50 metric (the mean average precision with IoU threshold of 0.5) calculates the average accuracy of object detection in the range from 0 to 1. The number of epochs started from 10 and changed in steps of 10 to 200 epochs. Model training time indicates how much time was spent training the model. The results of training models with different numbers of epochs and amounts of time are shown in Figure 1.

Based on the data obtained, we can conclude that the optimal number of epochs when training a model for yolov5 is 80, and for yolov8 is 60 epochs. Because for this number of epochs the model reaches mAP=0.702, and the training time is about 4 hours. Increasing the number of epochs only increases the training time of the model, but does not significantly increase its accuracy.

To measure the quality of models on the Internet, 10 videos were selected that contained one of the classes that was used to train the models - cigarettes. Each video was run through the models and the percentage of the appearance of a cigarette (Ap) in the frame and the detection of a cigarette by the model was calculated. Table 1 results.

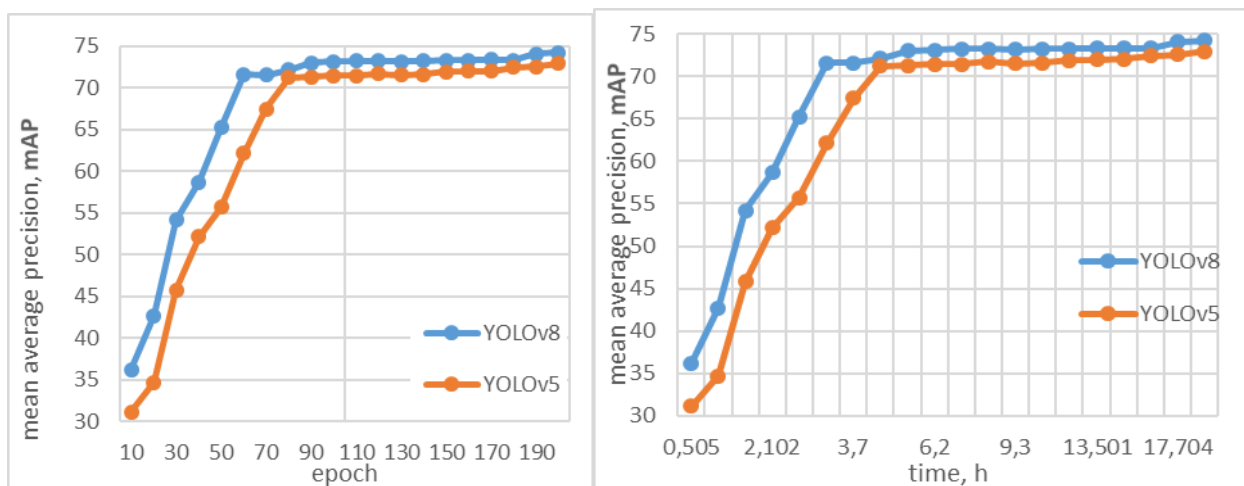


Figure 1 – The results of training models with different parameters

Table 1. Test results

Video number	Video quality, pixels	YOLOv5, Ap%	YOLOv8, Ap%
1	720	81	90
2	360	67	66
3	720	91	96
4	720	90	92
5	720	93	95
6	480	71	78
7	360	68	67
8	480	72	72
9	720	85	88
10	720	90	94

Based on the obtained data, it can be concluded that the quality of the models in terms of detecting objects in the video stream is approximately the same, but the difference is that the sizes of the original models are different. The analysis of the obtained results shows that it will be optimal to use a neural network version of yolo 8 in the software being developed, if the capabilities of the equipment on which the software will be run are weak.

V. DISCUSSION

The YOLOv8 neural network model is the best choice in terms of accuracy and speed, which is confirmed by experiment. Additionally, YOLOv8 goes from 0 to 1 quickly and is highly tuning, allowing hyperparameters to be tuned to further optimize its performance.

Also, the results of the experiment demonstrate that the size of the model and the training time of the model are important indicators, this is due to the fact that in the future the model will be used at state enterprises where there may be a shortage of capacity, so YOLOv8 also proved to be the optimal model for the given practical task.

VI. FUTURE RESEARCH

In the future, it is planned to develop software that will include a trained neural network YOLOv8 to identify violations of labor protection rules at the enterprise. In addition, it is planned to adjust the model more thoroughly and prepare data for training more thoroughly. Since the accuracy of detecting violations depends on the quality of the data.

VII. CONCLUSION

As part of this work, an analysis of the comparison of the architecture of two convolutional neural networks, namely YOLOv5 and YOLOv8, was carried out. Thanks to the analysis of these

models, it was established that they do not have a very large difference in the quality of the definition of the models, but it was possible to establish that the sizes of both models have a significant difference. This fact may affect the performance of the software using the trained model. Since the trained model can be used in software that will process video streams at different enterprises and inform the system about the violations that have occurred, the difference is significant. Therefore, the YOLOv8 convolutional neural network was chosen for more accurate object detection.

REFERENCES

- [1] Wang, Zijian & Wu, Yimin & Yang, Lichao & Thirunavukarasu, Arjun & Evison, Colin & Zhao, Yifan. (2021). Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*. 21. 3478. 10.3390/s21103478.
- [2] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, Bo Ma, A Review of Yolo Algorithm Developments, *Procedia Computer Science*, Volume 199, 2022, pp. 1066-1073, <https://doi.org/10.1016/j.procs.2022.01.135>.
- [3] Yonten Jamtsho, Panomkhawn Riyamongkol, Rattapoom Waranusast, Real-time license plate detection for non-helmeted motorcyclist using YOLO, *ICT Express*, Volume 7, Issue 1, 2021, pp. 104-109, <https://doi.org/10.1016/j.ict.2020.07.008>.
- [4] Yonten Jamtsho, Panomkhawn Riyamongkol, Rattapoom Waranusast, Real-time Bhutanese license plate localization using YOLO, *ICT Express*, Volume 6, Issue 2, 2020, pp. 121-124, <https://doi.org/10.1016/j.ict.2019.11.001>.
- [5] Chen R, Chen Y. Improved Convolutional Neural Network YOLOv5 for Underwater Target Detection Based on Autonomous Underwater Helicopter. *Journal of Marine Science and Engineering*. 2023; 11(5):989. <https://doi.org/10.3390/jmse11050989>
- [6] Bashir, Saba & Qureshi, Rizwan & Shah, Abbas & Fan, Xinqi & Alam, Tanvir. (2023). YOLOv5-M: A Deep Neural Network for Medical Object Detection in Real-time. 1-6. 10.1109/ISIEA58478.2023.10212322.
- [7] Zhai X, Huang Z, Li T, Liu H, Wang S. YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. *Electronics*. 2023; 12(17):3664. <https://doi.org/10.3390/electronics12173664>

Sentiment Analysis for Student Feedback

Halyna Melnyk, Vasyl Melnyk

*Faculty of Mathematics and Informatics
Yuriy Fedkovych Chernivtsi National University
Chernitsi, Ukraine*

Abstract— In this research paper, we investigate the efficacy of advanced NLP techniques, such as Long Short-Term Memory (LSTM) networks, Named Entity Recognition (NER), and Conditional Random Fields (CRF), for analyzing and interpreting high school students' feedback. Our focus is to refine educational strategies and classroom interaction by employing these cutting-edge methods.

Keywords—*Natural Language Processing, Sentiment Analysis, Software Tools.*

I. INTRODUCTION

Sentiment analysis in education is emerging as a pivotal tool in understanding and improving the learning environment. By harnessing the power of Natural Language Processing (NLP), educators can gain nuanced insights into students' emotional responses to teaching methods, course content, and overall classroom dynamics. This research explores the potential of sentiment analysis to interpret high school students' feedback, offering a data-driven approach to tailor educational strategies for enhanced student engagement and academic success.

II. RELATED WORK

Traditional methods of capturing student feedback have been reliant on manual reviews and observational techniques (Richardson, 2005). Recent studies have begun exploring the use of NLP in educational settings (Shaik, 2022). Our work builds upon these foundations, particularly focusing on the application of LSTM, NER, and CRF techniques in sentiment analysis (Gandhi, 2020).

III. METHODOLOGY

The research adopts a mixed-methods design to explore the application of Natural Language Processing (NLP) in high school educational settings, specifically assessing its impact on teacher instruction and student engagement.

Data were systematically collected from three distinct streams: student feedback forms, transcripts of chat communications during class discussions, and 'point of view' submissions within the LMS. Feedback forms and chat transcripts provided a wealth of data for sentiment analysis, capturing spontaneous student reactions and engagement levels. The 'point of view' forms, designed to prompt reflective and thoughtful responses, allowed for deeper insight into the students' perceptions of the teaching process. The dataset comprised 200 student feedback forms, 180 chat transcripts, and 500 'point of view' reactions, providing a diverse range of linguistic expressions.

Quantitative data from student feedback forms and chat transcripts were analyzed using NLP techniques to categorize sentiments into positive, negative, and neutral classifications. The LMS 'point of view' forms were qualitatively evaluated, with the textual data being coded for recurring themes related to the effectiveness and reception of NLP-supported teaching methods. Cross-analysis was conducted to triangulate findings from the quantitative sentiment analysis with the qualitative themes, providing a multifaceted understanding of the NLP's impact.

For extracting specific information from chat transcripts or feedback forms, like names, places, or subjects being discussed, Named Entity Recognition (NER) technique using CRF (Conditional Random Fields) algorithm was used to label patterns of words in a sentence that are significant. The next step should be training the Long Short-Term Memory (LSTM) model (a type of Recurrent

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

Neural Network, - RNN) on a labeled dataset of student feedback to classify the sentiment as positive, negative, or neutral effectively. The LSTM model used featured three layers with 128 nodes each, and the preprocessing involved tokenization and normalization of the text. Post-processing included categorizing sentiments into positive, negative, and neutral. Time complexity was evaluated, noting that the model processed each entry in approximately 0.5 seconds.

To validate the sentiment analysis, a subset of data from both student feedback forms and chat transcripts was independently assessed by a secondary researcher to determine inter-rater reliability.

Ethical guidelines were strictly adhered to, with all student data anonymized to protect privacy. Participants were informed of the study's scope and provided consent for their data to be used for research purposes, in compliance with institutional and educational standards for ethical research.

IV. RESULTS

The development and further deployment of the EduMood Tracker application within the classroom setting allowed for real-time sentiment analysis of student feedback and classroom chat discussions. The EduMood Tracker demonstrated an accuracy of 85%, with a precision of 80%, recall of 78%, and an F1-score of 79%. These results indicate a high correlation between the application's sentiment analysis and the students' engagement levels, providing educators with actionable insights to tailor their teaching methods for improved learning outcomes.

V. DISCUSSION

Our research recognizes limitations in the scope of data collected, which may not capture the full spectrum of student and teacher experiences with NLP tools. Our analysis reveals the potential of using advanced NLP techniques in educational settings. While the results are promising, the scope of data and potential biases in self-reporting highlight the need for a broader application to fully understand the impact of these tools in diverse educational environments.

VI. FUTURE RESEARCH

Future research will focus on enhancing the EduMood Tracker's algorithmic sophistication to distinguish between complex emotional states and to integrate predictive analytics for proactively shaping educational strategies.

VII. CONCLUSION

The use of the EduMood Tracker for sentiment analysis in high school classrooms offers a valuable tool for teachers to understand and improve the educational experience based on student feedback.

VIII. ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of scientists of the Department of Mathematical Modelling and Department of Applied Mathematics and Information Technologies of the Yuriy Fedkovych Chernivtsi National University for scientific guidance given in discussions and technical assistance helped in the actual research.

IX. DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

REFERENCES

- [1] Bird, S., "Natural Language Processing with Python," 1st. ed., O'Reilly, 2009, P.504.
- [2] Richardson, J.T.E., Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, 2005. 30(4): p. 387-415.
- [3] Shaik, T. Review of Trends and Challenges in Adopting NLP Methods for Education Feedback Analysis. *IEEE Access*, 2022. V.10: p. 56720-56739.
- [4] Gandhi, H. Extracting aspect terms using CRF and bi-LSTM models. *Procedia Computer Science*, 2020. 167(2020): p. 2486-2495

Обертальний криптоаналіз деяких функцій ускладнення ARX-криптосистем

Сергій Яковлев, Денис Кобець

*Кафедра математичних методів захисту інформації
НН ФТІ, КПІ ім. Ігоря Сікорського
Київ, Україна*

Анотація—У даній роботі розглядаються перетворення, які є аналогами множення на малі константи, але використовують тільки операції логіки. Такі перетворення є перспективними для побудови швидких та надійних ARX-криптосистем. Для усіх перетворень, які розглядалися, знайдено чисельні значення параметрів стійкості до обертального криптоаналізу.

Ключові слова—ARX-криптосистеми, Булеві функції, Обертальний криптоаналіз.

I. ВСТУП

ARX-криптосистеми будуються на основі виключно простих операцій, доступних на рівні інструкцій обчислювальних процесорів: модульного додавання, побітового додавання, циклічних зсувів та, у розширеному розумінні, інших операцій логіки (ТА, АБО, нециклічних зсувів тощо). Через просту реалізацію та надвисоку швидкість роботи такі криптосистеми стали важливою частиною так званої «легкої криптографії» (lightweight cryptography) – напрямку, присвяченому розробці надійних алгоритмів для малоресурсних пристроїв та Інтернету речей.

Алгебраїчна структура ARX-систем призвела до появи специфічного типу криптоаналізу, так званого обертального криптоаналізу (rotational cryptanalysis), який досліджує особливості зміни пар текстів, які відрізняються циклічним зсувом, під час виконання обчислень. Обертальний криптоаналіз дозволяє будувати ефективні атаки на ARX-системи, тому захищеність від нього наразі є однією з обов'язкових умов для надійних криптосистем даного типу.

У даній роботі розглядаються нелінійні перетворення спеціального типу, які імітують операцію множення на малу константу, але з використанням суто операцій логіки. Такі перетворення можуть використовуватись для підвищення швидкодії ARX-криптосистем при збереженні рівня стійкості. Для таких перетворень знайдено імовірності проходження пар обертання, що дозволить оцінювати стійкість до обертального криптоаналізу.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Обертальний криптоаналіз ARX-криптосистем був запропонований та розвинений Д. Ховратовичем та І. Ніколічем у роботах [1-3]. У даних роботах акцент робився, в першу чергу, на аналіз властивостей функції додавання за модулем як основного нелінійного перетворення ARX-криптосистем. На основі одержаних теоретичних результатів було проведено аналіз стійкості багатьох криптосистем даного типу.

Сучасні ARX-криптосистеми можуть використовувати окрім або замість додавання за модулем інші нелінійні перетворення. Так, розробники шифру NORX [4] запропонували спеціальну операцію, яка апроксимує додавання за модулем, однак використовує тільки операції логіки, що дозволило підвищити швидкість роботи даного шифру; у роботі [4] досліджено стійкість запропонованої апроксимуючої операції до обертального криптоаналізу. У криптосистемі Shabal використовуються такі перетворення, як множення на малі константи 3 та 5; такі перетворення виражаються через модульне додавання та операції

логіки і зберігають частину криптографічних властивостей модульного додавання. У роботі [5] Г. ван Ассе навів результати аналізу стійкості таких операцій до обертого криптоаналізу.

У даній роботі розглядаються перетворення, які моделюють операцію множення на константу 3, але із використанням тільки операцій логіки. Для таких перетворень буде проведено аналіз стійкості до обертого криптоаналізу.

III. ОСНОВНІ РЕЗУЛЬТАТИ

Нехай $V_n = \{0, 1\}^n$. Кожен елемент $x \in V_n$ розглядається як двійковий вектор та як двійковий запис певного натурального числа з інтервалу від 0 до $2^n - 1$:

$$x = (x_{n-1}, x_{n-2}, \dots, x_1, x_0) = x_0 + x_1 \cdot 2 + \dots + x_{n-1} \cdot 2^{n-1}.$$

Над двійковими векторами розглядаються логічні (або бітові) операції: логічне ТА (&), логічне АБО (v), стрілку Пірса (l), штрих Шеффера (|), імплікації (\rightarrow, \leftarrow), виключне АБО (\oplus) та еквівалентність (\sim); для логічного ТА будемо також використовувати спрощений запис xu замість $x \& u$. Заперечення бітів вектору x (інверсія) позначається як \bar{x} . Усі ці операції виконуються окремо над кожним бітом аргументів. Символами \gggg та \llll позначаються циклічні зсуви векторів, \gg та \ll – нециклічні зсуви векторів. Для вектору $x \in V_n$ позначимо його циклічний зсув на r позицій як $x^r = x \lll r$. Пару векторів (x, x^r) будемо називати *парою обертання*.

Основним параметром, який характеризує стійкість відображення $f: V_n \rightarrow V_n$ до обертого криптоаналізу, є імовірність проходження пари обертання через f :

$$rp^f(r) = \Pr_x \{f(x^r) = (f(x))^r\}.$$

Розглянемо функції виду $f(x) = x \star (x \ll 1)$, де \star позначає деяку бінарну операцію логіки. Такі функції є аналогами функції множення на малу константу 3:

$$3x \equiv x + 2x \equiv x + (x \ll 1) \pmod{2^n};$$

але замість відносно складної операції модульного додавання будуть використовуватись більш прості операції логіки.

Позначимо $u = f(x^r)$, $v = (f(x))^r$; тоді

$$\begin{array}{lll} i = 0 : & u_0 = x_r \star 0, & v_0 = x_r \star x_{r-1}; \\ 1 \leq i < n - r : & u_i = x_{r+1} \star x_{r+i-1}, & v_i = x_{r+1} \star x_{r+i-1}; \\ i = n - r : & u_{n-r} = x_0 \star x_{n-1}, & v_{n-r} = x_0 \star 0; \\ n - r < i < n : & u_i = x_{i-(n-r)} \star x_{i-(n-r)-1}, & v_i = x_{i-(n-r)} \star x_{i-(n-r)-1}. \end{array}$$

Бачимо, що вектори u та v можуть відрізнитись лише в двох позиціях: $i = 0$ та $i = n - r$; звідси випливає, що

$$rp^f(r) = \Pr_x \{u = v\} = \Pr_x \{x_r \star 0 = x_r \star x_{r-1}, x_0 \star 0 = x_0 \star x_{n-1}\}.$$

При $2 \leq r \leq n - 2$ наведені рівності незалежні, але при $r = 1$ та $r = n - 1$ вони використовують спільні біти. Відповідно, маємо:

$$\begin{aligned} rp^f(1) &= \Pr_x \{x_1 \star 0 = x_1 \star x_0, x_0 \star 0 = x_0 \star x_{n-1}\}, \\ rp^f(n - 1) &= \Pr_x \{x_{n-1} \star 0 = x_{n-1} \star x_{n-2}, x_0 \star 0 = x_0 \star x_{n-1}\}, \\ rp^f(r) &= \Pr_x \{x_r \star 0 = x_r \star x_{r-1}\} \cdot \Pr_x \{x_0 \star 0 = x_0 \star x_{n-1}\} \end{aligned}$$

для $2 \leq r \leq n - 2$. Зауважимо, що з наведених виразів випливає, що значення $rp^f(1)$ та $rp^f(n - 1)$ для заданого класу функцій завжди співпадають, а для $2 \leq r \leq n - 2$ справедлива рівність $rp^f(r) = p^2$, де $p := \Pr_x\{x_0 \star 0 = x_0 \star x_{n-1}\}$. Це спрощує обчислення імовірностей пар обертання.

Розглянемо $f(x) = x \& (x \ll 1)$. З наведених співвідношень маємо:

$$rp^f(1) = \Pr\{0 = x_0x_1, 0 = x_0x_{n-1}\}, \quad p = \Pr\{0 = x_0x_{n-1}\}.$$

Складемо таблицю істинності:

x_0	x_1	x_{n-1}	x_0x_1	x_0x_{n-1}
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	0	0
1	0	0	0	0
1	0	1	0	1
1	1	0	1	0
1	1	1	1	1

З таблиці безпосереднім обчисленням знаходимо $rp^f(1) = \frac{5}{8}$, $p = \frac{3}{4}$, звідки одержуємо $rp^f(r) = p^2 = \frac{9}{16}$ при $2 \leq r \leq n - 2$.

Аналогічно розглядаємо для побудови відображень виду $f(x) = x \star (x \ll 1)$ інші операції логіки та знаходимо значення імовірностей проходження пар обертання. Усі одержані результати наведено у таблиці 1.

ТАБЛИЦЯ 1. Імовірності проходження пар обертання через нелінійні перетворення виду $f(x) = x \star (x \ll 1)$ для усіх можливих значень зсуву векторів.

Функція	$rp^f(1), rp^f(n - 1)$	$rp^f(r), 2 \leq r \leq n - 2$
$f(x) = x \& (x \ll 1)$	$\frac{5}{8}$	$\frac{9}{16}$
$f(x) = x \vee (x \ll 1)$	$\frac{1}{2}$	$\frac{9}{16}$
$f(x) = x \downarrow (x \ll 1)$	$\frac{1}{2}$	$\frac{9}{16}$
$f(x) = x (x \ll 1)$	$\frac{5}{8}$	$\frac{9}{16}$
$f(x) = x \rightarrow (x \ll 1)$	$\frac{5}{8}$	$\frac{9}{16}$
$f(x) = x \leftarrow (x \ll 1)$	$\frac{1}{2}$	$\frac{9}{16}$
$f(x) = x \oplus (x \ll 1)$	$\frac{1}{4}$	$\frac{1}{4}$
$f(x) = x \sim (x \ll 1)$	$\frac{1}{4}$	$\frac{1}{4}$

IV. ОБГОВОРЕННЯ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Наведені у таблиці 1 результати показують, що усі перетворення, які розглядаються, можна розбити на два класи: нелінійні (з $\&$, \vee , \downarrow , $|$, \rightarrow , \leftarrow) та лінійні (з \oplus , \sim), у відповідності до алгебраїчних властивостей використаної операції. Імовірності проходження пари обертань через нелінійні перетворення складають не менше 0.5 і не більше 0.625; цікаво, що і найменші, і найбільші значення досягаються для найменших зсувів у парі обертань (на одну позицію ліворуч чи праворуч). Для лінійних перетворень імовірність проходження пари обертань завжди 0.25 для будь-яких значень зсувів. Окремо треба відзначити, що значення імовірностей пар обертань не залежать від довжини векторів, тому дані результати необмежено масштабуються.

У роботі [5] показано, що імовірності проходження пар обертань через функцію $f(x) = 3x \bmod 2^n$ залежать від n , але прямують до значення $1/3$. Відповідно, нелінійні аналоги даного відображення мають вищу імовірність проходження пар обертань, що можна розглядати як певну плату за більш просту та швидко обчислювану структуру. В той же час лінійні перетворення мають нижчу імовірність проходження пар обертань і, відповідно, надають більше стійкості до обертального криптоаналізу. Однак треба пам'ятати, що лінійні перетворення не надають стійкості до інших видів криптоаналізу, таких як диференціальний та лінійний. Відповідно, при побудові ARX-криптосистем із використанням аналогів множення на константи бажано комбінувати нелінійні та лінійні операції, щоб досягти потрібного рівня стійкості до усіх видів криптоаналізу.

Як напрямки подальших досліджень можна окреслити узагальнення одержаних результатів на інші види перетворень, зокрема, більш складної структури, та одержання оцінок стійкості до комбінованих методів криптоаналізу, таких як диференціально-обертальний криптоаналіз.

V. ВИСНОВКИ

У даній роботі розглянуто перетворення, які є аналогами множення на константу 3, але використовують тільки операції логіки. Такі перетворення є перспективними для побудови швидких та захищених ARX-криптосистем. Для даних перетворень було знайдено значення параметрів, які характеризують стійкість до обертального криптоаналізу. Показано, що дані значення не залежать від довжини вектору і для нелінійних перетворень, які розглядалися, поступаються традиційному множенню на константу 3, однак для лінійних перетворень є кращими. Одержані результати дозволять проводити аналіз та будувати ефективні за швидкістю і захищені від обертального криптоаналізу ARX-криптосистеми.

VI. ПОДЯКИ

Автори вдячні Максиму Недождю, бакалавру НН ФТІ КПІ ім. Ігоря Сікорського, за допомогу у підготовці даної доповіді до публікації.

ЛІТЕРАТУРА

- [1] D. Khovratovich, and I. Nikolić, "Rotational Cryptanalysis of ARX," in *Fast Software Encryption 2010*, Lecture Notes in Computer Science, vol. 6147, pp. 333–346, 2010 — DOI: https://doi.org/10.1007/978-3-642-13858-4_1
- [2] D. Khovratovich, I. Nikolić, J. Pieprzyk, P. Sokolowski, and R. Steinfield, "Rotational Cryptanalysis of ARX Revisited", *Cryptology ePrint Archive*, Paper 2015/095, 2015, <https://eprint.iacr.org/2015/095>
- [3] D. Khovratovich, I. Nikolić, and C. Rechberger, "Rotational rebound attacks on reduced Skein," in *Advances in Cryptology – ASIACRYPT 2010*, Lecture Notes in Computer Science, vol. 6477, pp. 1-19, 2010. — DOI: https://doi.org/10.1007/978-3-642-17373-8_1
- [4] J.P. Aumasson, P. Jovanovic, and S. Neves, "Analysis of NORX: Investigating differential and rotational properties," *Cryptology ePrint Archive*, Report 2014/317, 2014, <http://eprint.iacr.org/2014/317>
- [5] G. V. Assche, "A rotational distinguisher on Shabal's keyed permutation and its impact on the security proofs", 2010, available online at <http://gva.noekeon.org/papers/ShabalRotation.pdf>.

Підходи до автентифікованого шифрування

Наталія Щур

*Кафедра комп'ютерної інженерії та кібербезпеки
Державний університет «Житомирська політехніка»
Житомир, Україна*

Анотація—У роботі розглянуто базові концепції та принципи, що лежать в основі автентифікованого шифрування. Наведено приклади відомих алгоритмів, які одночасно забезпечують як конфіденційність даних, так і їх цілісність. Проаналізовано основні підходи до автентифікованого шифрування, описано переваги та недоліки кожного підходу. Представлено програмну реалізацію автентифікованого шифрування за алгоритмом AES-GCM.

Ключові слова—*симетричне шифрування, автентифіковане шифрування, AEAD.*

I. ВСТУП

Шифрування відіграє важливу роль у забезпеченні конфіденційності даних, які передаються через мережу Інтернет або зберігаються в комп'ютерних системах. За допомогою правил шифру (алгоритму шифрування) і секретного ключа, повідомлення (відкритий текст) перетворюються у зашифрований вигляд (шифротекст). Хоча шифрування призначене для того, щоб неавторизовані особи не могли прочитати отримані ними дані, у деяких випадках зломисник потенційно може змінити зашифрований текст або підробити його без знання ключа. До того ж, іноді отримувач не може бути впевненим, що повідомлення було надіслане саме відправником і не було підроблене або змінене при передачі. Враховуючи ці недоліки, автентифіковане шифрування (Authenticated Encryption, AE) стало широко використовуваним методом, який одночасно забезпечує конфіденційність даних, їх цілісність та гарантує автентичність.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Автентифіковане шифрування є актуальною темою дослідження в галузі криптографії, яка вивчає методи захисту інформації від несанкціонованого доступу, зміни або підробки. Багато науковців з різних країн та організацій присвятили свої роботи розробці, аналізу та застосуванню алгоритмів автентифікованого шифрування для різних сценаріїв і потреб.

Значний внесок у дослідження основних концепцій та алгоритмів автентифікованого шифрування зробили такі вчені як J.Black [1], P.Rogaway [2], M.Bellare [3] та інші. Автор статті [2] запропонував перші практичні схеми автентифікованого шифрування із присланими даними. У роботі [4] представлено загальну класифікацію та огляд автентифікованого шифрування, його основних характеристик, параметрів безпеки та функціональних властивостей. У статті [5] розглядаються стандартні алгоритми AE, недоліки класичних моделей безпеки для алгоритмів AE та можливі атаки.

III. ПРИНЦИПИ АВТЕНТИФІКОВАНОГО ШИФРУВАННЯ

Автентифіковане шифрування комбінує як шифрування, так і автентифікацію в одному криптографічному алгоритмі. Для перевірки автентичності даних використовується код автентифікації повідомлення (Message Authentication Code, MAC), який також називають тегом автентифікації. Обчислення кодів автентифікації зазвичай відбувається за допомогою криптографічних функцій хешування, наприклад, HMAC (Hash-based Message Authentication Code). MAC є унікальним значенням фіксованого розміру, яке забезпечує цілісність повідомлення, оскільки будь-яка зміна повідомлення призведе до іншого значення MAC.

Код автентифікації додається до зашифрованого повідомлення і служить для перевірки його справжності та цілісності при дешифруванні. Якщо код автентифікації не співпадає з очікуваним, це означає, що повідомлення було змінено або підроблено, і тому воно відхиляється.

Для правильної роботи АЕ також часто використовують *nonce* – унікальне значення, яке застосовується один раз для кожного повідомлення. *Nonce* служить для запобігання атакам або аналізу, коли зловмисник намагається використати одне і те ж зашифроване повідомлення кілька разів або порівняти різні зашифровані повідомлення з однаковим ключем.

Прикладами АЕ алгоритмів є:

- AES-GCM: складається з блокового шифру AES у режимі лічильника (CTR, Counter Mode) для шифрування та функції хешування GHASH (Galois Hash) для обчислення тегу автентифікації. GCM є однією з найпопулярніших схем автентифікованого шифрування, що стандартизована NIST і використовується в протоколах TLS, IPSec, SSH тощо.
- AES-CCM: поєднує блокове шифрування в режимі лічильника з автентифікацією CBC-MAC. AES-CCM часто використовується для бездротових мереж і IoT пристроїв, але є менш ефективним за AES-GCM.
- ChaCha20-Poly1305: потоковий шифр ChaCha20 з функцією Poly1305 для обчислення MAC. ChaCha20 приблизно втричі швидший за AES на платформах, які не мають апаратних криптоприскорювачів AES.

IV. РЕЗУЛЬТАТИ

Ефективність автентифікованого шифрування залежить від правильного вибору методів та їхнього використання для захисту даних в різних ситуаціях. Існує кілька підходів до автентифікованого шифрування і вибір конкретного варіанта залежить від потреб безпеки, продуктивності та вимог до платформи або протоколу, у якому він буде використовуватись.

Зашифрування, потім обчислення MAC (Encrypt-then-MAC, EtM). Відкритий текст спочатку шифрується із застосуванням секретного ключа та алгоритму шифрування. На основі отриманого зашифрованого тексту обчислюється MAC (рис. 1). Зауважимо, що для шифрування та хешування (обчислення MAC) використовуються різні ключі. Зашифрований текст та його MAC надсилаються разом. Це широко використовуваний підхід, який забезпечує чітке розмежування між процесами шифрування та автентифікації.

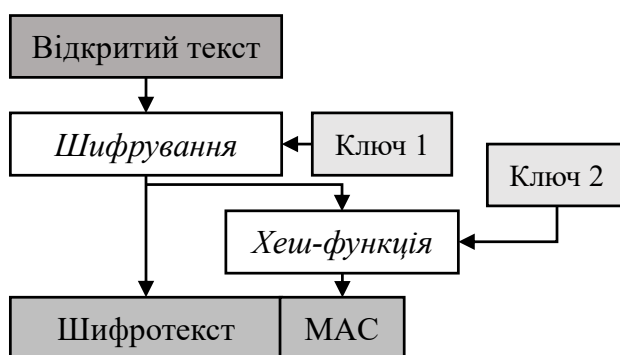


Рис. 1. Схема автентифікованого шифрування згідно підходу *EtM*

Зашифрування та обчислення MAC (Encrypt-and-MAC, E&M). MAC обчислюється на основі відкритого тексту. У свою чергу відкритий текст шифрується окремо без MAC. MAC відкритого тексту та зашифрований текст надсилаються разом. Для шифрування та хешування використовується один і той самий ключ (рис. 2). Хоча цей підхід менш поширений через потенційну вразливість, він може бути безпечним, якщо його застосовувати обережно.

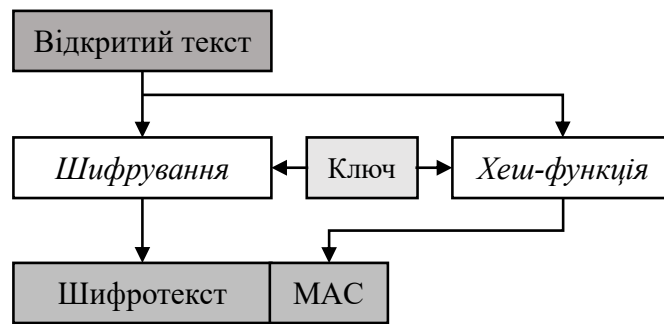


Рис. 2. Схема автентифікованого шифрування згідно підходу E&M

Обчислення MAC, потім зашифрування (MAC-then-Encrypt, MtE). На основі відкритого тексту обчислюється MAC, після чого відкритий текст і MAC об'єднуються і вже разом шифруються симетричним алгоритмом шифрування (рис. 3). Ключ шифрування може збігатися або не збігатися з ключем, що використовується для створення MAC. Цей підхід вважається менш безпечним і не рекомендується, оскільки він може призвести до витоку інформації про відкритий текст.

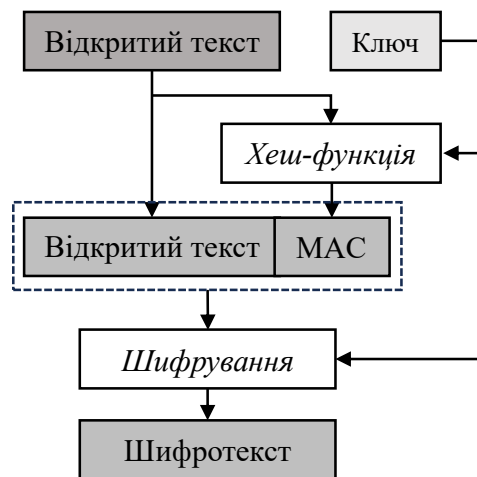


Рис. 3. Схема автентифікованого шифрування згідно підходу MtE

Порівняння підходів до автентифікованого шифрування, їх переваги та недоліки наведено у табл. 1.

Таблиця 1

Переваги та недоліки підходів до автентифікованого шифрування

Назва підходу	Переваги	Недоліки
<i>Encrypt-then-MAC</i>	Автентифікація виконується на основі шифротексту, тому будь-яка помилка автентифікації може бути виявлена до розшифрування повідомлення	Вимагає два ключі (один для шифрування, інший для автентифікації), що може призвести до зниження продуктивності
<i>Encrypt-and-MAC</i>	Паралельне виконання операцій шифрування та автентифікації	Автентифікація виконується на основі відкритого тексту, що може призвести до атаки на цілісність
<i>MAC-then-Encrypt</i>	Дозволяє застосувати автентифікацію до чітко визначеної точки даних перед шифруванням	Нестійкий до багатьох атак, зокрема до атак бічними каналами

Розглянемо найпростіший варіант реалізації підходу Encrypt-then-MAC для оптимального поєднання шифрування та обчислення MAC (рис.4). Ми будемо використовувати мову програмування Python та бібліотеку `pycryptodome`, яка підтримує алгоритм автентифікованого шифрування AES-GCM. Для встановлення бібліотеки потрібно відкрити командний рядок та ввести команду `pip install pycryptodome`.

Функція `encrypt_AES_GCM(msg, secretKey)` шифрує повідомлення, використовуючи ключ шифрування, та повертає шифротекст, унікальний попсе (використовується тільки один раз для конкретного повідомлення) та MAC. В режимі GCM тег автентифікації (MAC) обчислюється на основі шифротексту.

Функція `decrypt_AES_GCM(encryptedMsg, secretKey)` дешифрує шифротекст, використовуючи ключ шифрування, попсе та MAC, та повертає відкритий текст.

```
from Crypto.Cipher import AES
import binascii, os

def encrypt_AES_GCM(msg, secretKey):
    aesCipher = AES.new(secretKey, AES.MODE_GCM)
    ciphertext, authTag = aesCipher.encrypt_and_digest(msg)
    return (ciphertext, aesCipher.nonce, authTag)

def decrypt_AES_GCM(encryptedMsg, secretKey):
    (ciphertext, nonce, authTag) = encryptedMsg
    aesCipher = AES.new(secretKey, AES.MODE_GCM, nonce)
    plaintext = aesCipher.decrypt_and_verify(ciphertext, authTag)
    return plaintext

secretKey = os.urandom(32) # 256-бітовий випадковий ключ
print("\nКлюч шифрування:", binascii.hexlify(secretKey).decode('utf-8'))

msg = b'Hello world'
encryptedMsg = encrypt_AES_GCM(msg, secretKey)
print("\n=====Шифрування:=====")
print("Відкритий текст:", msg.decode('utf-8'))
print("Шифротекст:", binascii.hexlify(encryptedMsg[0]).decode('utf-8'))
print("nonce:", binascii.hexlify(encryptedMsg[1]).decode('utf-8'))
print("MAC:", binascii.hexlify(encryptedMsg[2]).decode('utf-8'))

decryptedMsg = decrypt_AES_GCM(encryptedMsg, secretKey)
print("\n=====Дешифрування:=====")
print("Plaintext:", decryptedMsg.decode('utf-8'))
```

Рис. 4. Автентифіковане шифрування за алгоритмом AES-GCM

Результат роботи алгоритму автентифікованого шифрування AES-GCM наведено нижче (рис.5). Як видно, програма генерує 256-бітовий випадковий ключ шифрування за допомогою функції `os.urandom(32)`.

```
Ключ шифрування: 7a9685dd1a4b56f105ef341c6ae6f63579028ebff5536c1f59c2f4bbdb9c3a6

=====Шифрування:=====
Відкритий текст: Hello world
Шифротекст: 4d5346021abab5353f25eb
nonce: 691683db9ee12d6cd89332d8ae66fa19
MAC: 9836cc2efd644cd484684a7f6fe37e6d

=====Дешифрування:=====
Plaintext: Hello world
```

Рис. 5. Результат роботи алгоритму автентифікованого шифрування AES-GCM

V. ОБГОВОРЕННЯ

Найбільш безпечним підходом до автентифікованого шифрування вважається Encrypt-then-MAC. Він гарантує цілісність шифротексту, оскільки MAC перевіряється перед розшифруванням. Цей підхід забезпечує конфіденційність відкритого тексту, оскільки

шифрування виконується перед MAC. Encrypt-then-MAC дозволяє використовувати будь-який алгоритм шифрування і будь-який алгоритм MAC, тому що вони працюють незалежно один від одного. Це забезпечує гнучкість і сумісність різних алгоритмів.

Також, на сьогоднішній день все більшої популярності набуває *автентифіковане шифрування з приєднаними даними* (Authenticated Encryption with Associated Data, AEAD), яке дозволяє автентифікувати не лише саме повідомлення, але й додаткові дані, які не підлягають шифруванню. Це корисно в багатьох контекстах, наприклад, у разі використання AEAD для захисту мережевого пакета. Якщо дані пакетуються та надсилаються через мережу, пакет часто містить заголовок із метаданими, такими як адреси відправника та одержувача. Ці дані не можуть бути зашифровані під час надсилання, оскільки інші вузли на маршруті повинні мати можливість прочитати цю інформацію, щоб пакет можна було надіслати правильному пункту призначення. Таким чином приєднані дані використовуються для захисту інформації, яка потребує автентифікації, але не має бути конфіденційною.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Автентифіковане шифрування залишається ключовим компонентом для забезпечення конфіденційності та цілісності інформації в сучасному цифровому світі, вимагаючи постійного вивчення, вдосконалення та пристосування до нових викликів та можливостей. Перспективи подальших досліджень полягатимуть у огляді застосувань автентифікованого шифрування у нових областях, зокрема таких як легковагова криптографія.

VII. ВИСНОВКИ

Вибір підходу до автентифікованого шифрування залежить від конкретних вимог програми, базових криптографічних примітивів і моделі загрози. Важливо дотримуватися встановлених криптографічних стандартів і найкращих практик під час впровадження автентифікованого шифрування, щоб забезпечити його безпеку та продуктивність.

ЛІТЕРАТУРА

- [1] J. Black. "Authenticated Encryption. Encyclopedia of Cryptography and Security". Boston, MA, 2011. pp. 52–61.
- [2] P. Rogaway. "Authenticated-encryption with associated-data". *Proceedings of the 9th ACM Conference on Computer and Communications Security*, Washington, DC, USA. 2002, pp. 98-107.
- [3] M. Bellare, C. Namprempre. "Authenticated encryption: Relations among notions and analysis of the generic composition paradigm". *International Conference on the Theory and Application of Cryptology and Information Security*. Berlin, Heidelberg: Springer Berlin Heidelberg. 2000. pp. 531-545.
- [4] M. A. Jimale et al. "Authenticated Encryption Schemes: A Systematic Review" in *IEEE Access*. 2022, vol. 10, pp 14739-14766.
- [5] D. Maimut and R. Reyhanitabar. "Authenticated Encryption: Toward Next-Generation Algorithms" in *IEEE Security & Privacy*, vol. 12, no. 2, pp. 70-72.

Real-time big data analysis systems resulting from the Internet of Things IoT.

Mohammed A. Makarem
Department of Computer Engineering
King Fahd university of petroleum and
minerals
Dhahran, Saudi Arabia
Makarem.a.Mohammed@gmail.com

Muneef A. Razaz
Department of Electrical Engineering
King Fahd university of petroleum and
minerals
Dhahran, Saudi Arabia
mneefbasha@gmail.com

Abstract— With the development of advanced telecommunications, the widespread availability of high-speed networks, smartphones, and smart devices, the term "Internet of Things" (IoT) has emerged along with the importance of data generated by it. This importance has increased significantly with the notable increase in the quantity of data, now commonly referred to as "big data". Various studies have emerged that seek to utilize IoT data to drive development in various fields, thus necessitating frameworks that combine big data technologies and IoT services. One of the challenges of dealing with IoT data is its fast flow and the need to collect and process it in real-time. Suitable big data frameworks have been studied to address this challenge. This research compared the Hadoop and Spark systems to select the most appropriate framework for IoT data that meets the requirements for fast flow and analysis, as well as ease of use. The results showed that Spark outperforms Hadoop, particularly in its ability to enable real-time data analysis, speed of processing, and efficient memory usage.

Keywords— *IoT (Internet of Things)-Big Data - Hadoop – Spark.*

I. INTRODUCTION

Big Data (the new oil)" is one of the most common terms nowadays [1], as it sheds light on the importance of data in a world where digital information is present and ingrained in all aspects of our lives and society. It appears that the growth in information production cannot be stopped, as companies continue to gather data such as sales data, operational data, financial data, human resource data, consumer data, and more, with the aim of analyzing it, gaining specific insights, and using it to make decisions. Social media platforms are one of the main sources of data. Every day, 500 million new tweets, 294 billion emails, 4 million gigabytes of Facebook data, 65 billion WhatsApp messages, and 720,000 hours of new content are added to YouTube [2]. All of this is without mentioning the spread of IoT devices that has led to an explosion of digital data that was previously not available. The quantity of data that is generated, collected, duplicated, and utilized worldwide is expanding at a rapid pace [3]. By the year 2020, the size of data had reached 64.2 zettabytes, marking a new threshold. This growth surpassed previous projections, primarily due to the COVID-19 pandemic, which caused a surge in demand as more individuals began working and learning from home, as well as increasing their use of home entertainment services.

It is expected that the growth and volume of this data will continue to increase, reaching over 180 zettabytes over the next four years until 2025 (see Figure 1).

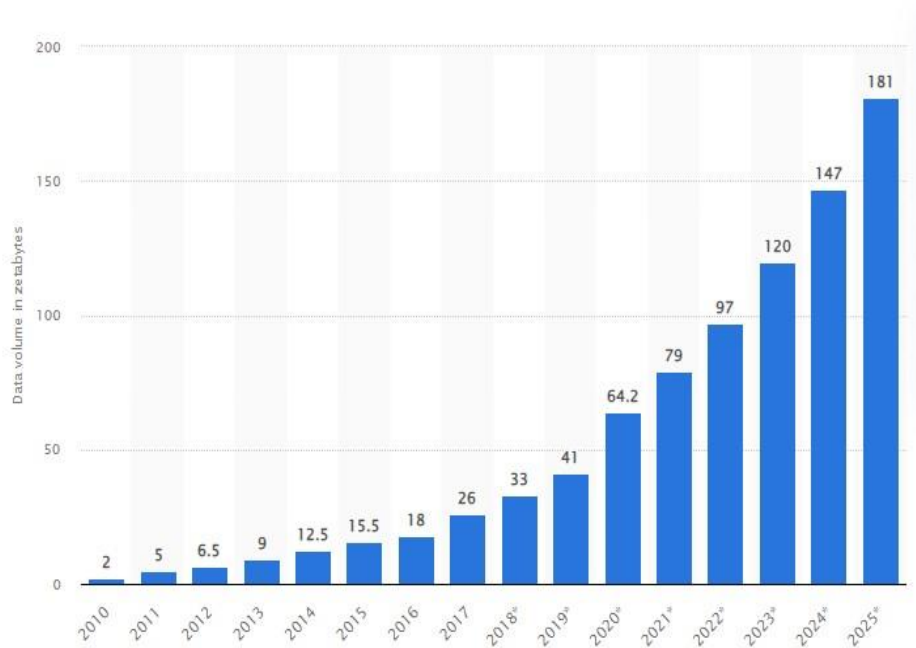


Figure 1. Data production from 2010 to 2025 in zettabytes [23].

Since data is information in a raw form, it was necessary to use scientific methods, processing techniques, algorithms, and systems to extract the knowledge and ideas necessary for development in all fields. This was previously done using traditional storage and display methods that are no longer suitable due to the enormous increase in the amount of data produced and its diverse sources, which has led to the term "Big Data". Specialized platforms are now required to deal with this data (storage, analysis, and representation) in order to obtain useful results. In recent years, the IoT has exploded, contributing to the massive increase in the quantity and diversity of data. Any physical object can be easily transformed into a part of the IoT by connecting it to a sensor device [2].

It is expected that the amount of data generated by IoT devices will reach 79.4 zettabytes of data by the year 2025 [4]. Additionally, the number of IoT devices worldwide is also expected to nearly triple from 8.74 billion in 2020 to over 25.4 billion devices by the year 2030 (see Figure 2). Not only IoT networks will be in the future, but 5G and 6G networks will generate hybrid data and needs what is called network slicing concept [3]. By 2025, 30% of all data will be real-time data, with IoT representing nearly 95% of it, and 20% of all data will be critical, with 10% being highly important [4].

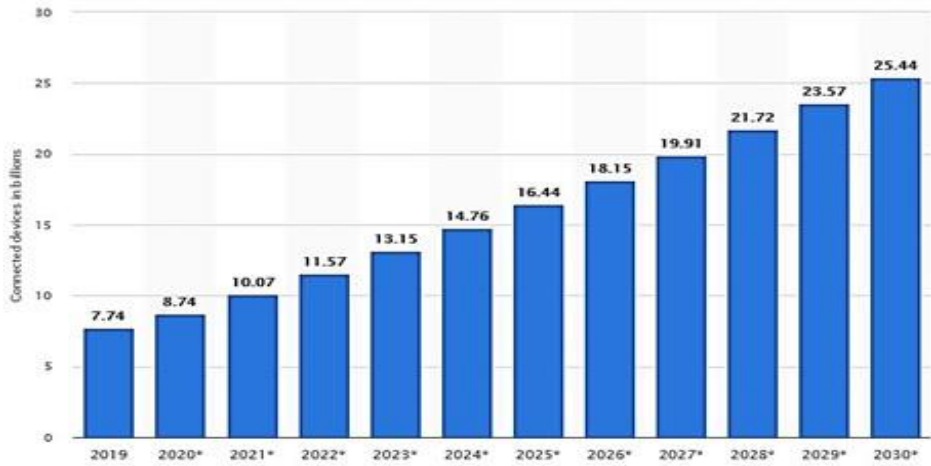


Figure 2 . number of connected IoT devices from 2019 to 2030 [23].

The data generated by IoT devices is considered more diverse and voluminous compared to other types of data, as it can be connected to any physical device. Therefore, IoT data is diverse, precise, and flows rapidly, resulting in a vast amount of data that requires analysis. IoT devices are used in various industry sectors and consumer markets, such as major industries like electricity, gas, steam, air conditioning, water supply, waste management, as well as in retail, wholesale, transportation, storage, and government. For example, smart buildings can collect data related to environmental conditions, such as air quality, pollution levels, temperature, humidity, and brightness, to determine what needs to be changed for human safety and comfort. They can also monitor energy usage patterns to determine when and how a building uses energy and take steps to improve energy efficiency. Additionally, they can monitor water usage and use information for predictive maintenance of building equipment. The collection and analysis of IoT data is facilitated by many IoT platforms, which use machine learning to analyze the data. For instance, sensor data that measures equipment vibrations and other factors can help detect anomalies and predict issues before serious problems arise. The ability to use all of this data results in more actionable insights and a greater return on investment. IoT devices, especially IoT sensors that capture valuable data, generate high-speed massive data streams that require processing and analysis in real-time. This involves monitoring values and detecting anomalies, and using them for mathematical and aggregation operations, among other use cases that help gain insights and understand reality faster. Since the value of IoT in smart environments comes from the knowledge derived from analyzing the data it provides, we need to analyze this data efficiently to support real-time processing operations, achieve maximum benefits, and obtain the best results quickly. This has led to an urgent need for frameworks that combine big data technologies and IoT services.

A. *Research problem*

As the number of IoT devices, including sensors and others, increases in smart environments, the amount of data flowing also increases, with different types that cannot be dealt with using traditional methods. The main challenge of IoT data is its real-time nature, requiring processing and analysis in real-time. Companies must perform real-time analytics to benefit from these types of data. There is a need for new techniques to process this data and to present the results more efficiently. Hence, this study raises the following question: What is the effectiveness of current data analysis frameworks in analyzing massive IoT data in real-time?

B. *research goals*

The purpose of this research is to investigate data analysis systems and evaluate their ability and efficiency in analyzing large amounts of real-time data generated by the IoT, which is constantly increasing with the growing number of network-connected devices, as well as to improve the monitoring of results and extracting information from IoT. This is accomplished by researching current data analysis systems, evaluating their efficiency in analyzing large amounts of IoT data, and determining their ability and suitability to analyses real-time IoT data.

II. USED RESEARCH PRINCIPLES

A. *IoT and Smart Environment*

The IoT is a system of computing devices, mechanical and digital machines, objects, animals, or people that are interconnected and have the ability to transfer data over a network without the need for human interaction. It provides unique identifiers (IP) to connect devices to the internet and interact with the external environment[5]. The thing can be a person with a heart monitor sensor or a car with an integrated pressure sensor to alert the driver to low tire pressure. Therefore, the thing is any device, terminal, or similar that can

be defined on the internet through an IP address and helps interact with the external environment. The concept of a smart environment is based on the initial idea of computing everywhere, promoting the idea of a world filled with sensors and devices connected to each other and integrated into our daily lives through a permanent and continuous network." The definition of a smart environment is also characterized as a small world where different types of smart devices work together to make human life more comfortable[6]. Figure 3 depicts some of the applications of smart environments that have spread widely due to their advanced features, such as remote-control devices for communication, interconnecting devices using middleware, gathering information through sensors, and transmitting it, and predictive capabilities and the ability to make decisions[7].

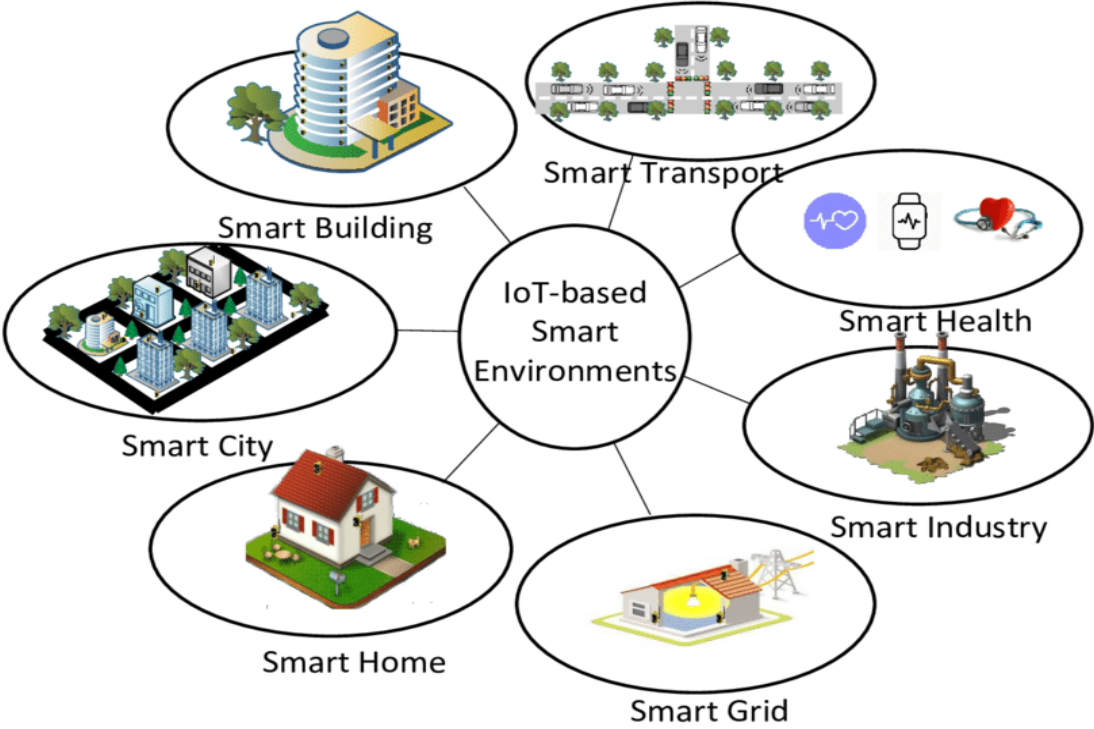


Figure 3. Smart environment applications [24].

B. Big Data

Big data refers to a large set of data that is difficult to store on a single device due to its size or complexity. It requires a cluster of devices (called clusters) and is challenging to process and quickly access using conventional tools or applications for data management[8]. Data flows at a high rate (meaning the rate at which it is generated and changed), and it must be processed with minimal response time. Big data requires special and advanced technologies and new forms of processing to analyze it and extract results. Its importance lies in the added value of understanding it, as it enhances the decision-making process and deepens understanding in various fields.

III. FRAMEWORK FOR BIG DATA PROCESSING SPARK AND HADOOP

Today, there are many frameworks for processing big data such as Apache Hadoop, Apache Cassandra, Apache Hive, Apache Spark, and others. The most common and widely used ones are Apache Spark and Apache Hadoop frameworks. Both are used in various studies that deal with processing big data in general and big data resulting from the IoT in particular. As they are our primary frameworks, we always question which framework to use. Therefore, we will define each framework and its components and make a comparison to determine which framework, Spark or Hadoop, is more suitable for processing data resulting from the IoT [9][10].

1. *Hadoop*

Apache Hadoop is a powerful framework that facilitates the distributed processing of large datasets across a cluster of machines, utilizing the MapReduce algorithm. This method separates the data into blocks and distributes them to various nodes within the cluster. This facilitates the processing of data in parallel on each node, which ultimately leads to distinct outputs from the algorithm. Hadoop stores data on every machine in the cluster and manages the files through the Hadoop Distributed File System (HDFS), which provides a scalable series of devices that can start with a single machine and expand to thousands, including any type of device produced by different companies. To ensure fault tolerance, Hadoop is equipped with a system that handles application-layer failure cases by copying data across the cluster when a device fails. This mechanism enables the framework to rebuild the missing component from another location, thereby preventing data loss.

The Apache Hadoop project consists of four primary units, each designed to fulfill a unique purpose. One of these units, HDFS, is responsible for managing the storage of large data sets across the cluster. HDFS is capable of handling both structured and unstructured data, as well as various storage devices, ranging from individual hard drives to enterprise-level drives. Hadoop's MapReduce component divides data chunks across the cluster, allowing for parallel processing by the nodes in the cluster, which then combine the results to produce the desired output. Yet Another Resource Negotiator, or YARN for short, is another component that oversees resource management, computation, and the scheduling of jobs. In conclusion, the Hadoop Common library provides a collection of tools and libraries that are designed for public use in order to support all of the Hadoop components. One of these fundamental components is the Hadoop core, which is also provided by the Hadoop common library.

2. *Spark*

Apache Spark is an open-source tool for processing data that can be utilized in a standalone mode or on cloud infrastructure. It is specifically designed for high-performance computing, utilizing in-memory random access storage to efficiently store and process large data sets. Spark supports various types of big data workloads, such as batch processing similar to MapReduce, real-time stream processing, machine learning, graph processing, and interactive queries. Spark offers easy-to-use and high-level APIs, which enable seamless integration with various libraries, including PyTorch and TensorFlow. The MapReduce algorithm provided the inspiration for the Spark engine, which was designed to increase the algorithm's effectiveness while preserving its advantages. Spark is able to access data that is stored on many storage systems despite the fact that it does not have its own file system. Spark's fundamental data structure is referred to as Resilient Distributed Datasets (RDDs for short) (RDD). The following are the five primary parts that make up Apache Spark:

Apache Spark is a distributed computing framework consisting of various components. The Spark Core is the foundation of the project, providing essential functions such as task dispatching, scheduling, error

recovery, and input/output operations. Spark Streaming is a component that facilitates real-time data processing, enabling data to be generated from various sources. Spark SQL is used for structured data processing and collects information about organized data and its processing requirements. MLlib is a library for machine learning that provides a variety of algorithms that were developed with the goal of making machine learning more approachable and scalable. Lastly, GraphX is a collection of application programming interfaces that are used for graph analytics jobs.

IV. THE PRIMARY DISTINCTIONS BETWEEN HADOOP AND SPARK

The following sections clarify the main similarities and differences between the two frameworks. We will examine Spark vs Hadoop from various angles such as performance, cost, and ease of use to find the most suitable framework for handling IoT data[11]. The Hadoop framework operates by accessing data that is stored on the HDFS hard disk, while Spark utilizes in-memory processing. According to Apache, Spark is significantly faster than Hadoop with MapReduce, performing 100 times faster when using Random Access Memory (RAM) for computation. Spark requires less hardware and can process 100 terabytes of data on HDFS ten times faster than Hadoop. The primary reason for Spark's superiority is that it uses RAM instead of reading and writing intermediate data to disk like Hadoop, making it more efficient in handling fast-producing IoT data that require speedy processing.

Both systems may be utilized without cost, but in order to evaluate the overall cost, it is necessary to take into account the expenses associated with the infrastructure, the maintenance, and the development of the system. The underlying hardware that is necessary to operate these products is the most important consideration when it comes to the cost category. Spark's costs are greater because it relies on in-memory computing for real-time data processing, but Hadoop's operational costs are very cheap since it can employ any form of disc storage for data processing. In contrast, Hadoop's cost of operation is quite low.

While both Hadoop and Spark are capable of managing data in a distributed setting, Hadoop is better suited for processing data in batches, whereas real-time processing is where Spark really shines. In contrast to MapReduce, Hadoop stores data in clusters and performs analysis on it in parallel, although MapReduce does not need a significant amount of RAM to process massive amounts of data. Spark is able to effectively manage direct data streams because of its utilization of in-memory calculations and high-level application programming interfaces. This differentiates Spark from other data processing frameworks, particularly when working with data from Internet of Things devices.

Scalability is where Hadoop and Spark differ, with Hadoop using the Hadoop Distributed File System (HDFS) to handle massive amounts of data, while Spark relies on the HDFS when dealing with extremely large amounts of data. Clusters in both frameworks can scale up by adding more nodes to the network, with no fixed limit on the number of nodes or the amount of data that can be processed. Scala, which is Spark's native language, along with Java, Spark SQL, R, and Python are among the various programming languages that are supported by Spark. In comparison, Hadoop mostly uses Java or Python for developing MapReduce code. Spark shell may be used to do data analysis in an interactive manner using either Scala or Python. This enables users to get immediate feedback on their queries and makes Spark more user-friendly than Hadoop. In addition, programmers can reuse previously written code when it is required, which cuts down on the amount of time required to build apps. Integration of historical and flowing data is also possible, which makes this process more efficient overall, particularly when used to applications for the Internet of Things.

Machine learning algorithms operate better with computational calculations in memory, making Spark a faster solution for this field. When workloads are broken up into parallel processes using the Hadoop MapReduce framework, the resulting processes may sometimes be too large for machine learning algorithms, which can lead to worries about the framework's input and output performance. Spark comes with the virtual machine learning library MLlib, which is empowered with computational calculations in memory and includes tools for regression, classification, constructing pipelines, evaluating, and much more. Spark, together with MLlib, is the preferred choice for machine learning and gives it an advantage in dealing with IoT data, which is often used for classification and prediction operations.

1. *Hadoop vs. Spark use cases*

Based on the comparison between Hadoop and Spark in the aforementioned domains, we can extract some usage scenarios for each framework.

Use cases for Hadoop include building the infrastructure for data analysis on a limited budget, completing jobs that do not require immediate results and where time is not a constraint, processing batches of data while leveraging disk I/O operations, and analyzing historical and archival data.

The following usage scenario can be separated, where Spark outperforms Hadoop:

Real-time streaming data analysis. When time is critical, Spark provides fast results with in-memory computational operations, handling parallel workflows using algorithmic techniques, parallel processing of graphical data for data modeling, and all machine learning applications.

V. LITERATURE REVIEW

The purpose of this literature review is to explain the importance and methods of analyzing big data in various fields and different types of data. The platforms used for analysis vary, with Hadoop and Spark being the most common. The sources of big data also vary, including data generated by companies, customer transactions, banking transactions, social media, and others in general, as well as data generated by IoT devices in particular. This section sheds light on a set of studies that were classified into studies that used data from public sources and studies that specialized in the field of IoT, along with the platforms used for analysis.

A. *general data*

Retail companies understand the need for analyzing and predicting their sales and customer behavior towards their products and product categories. With the assistance of big data frameworks, retail companies can create customized deals and promotional offers for their customers, allowing them to handle large sales volumes more efficiently. Apache Spark framework was used to analyze Black Friday[12] sales data, and machine learning models were trained using the built-in MLlib library to predict future prices and sales. Linear regression[13] and Random Forest models[14] were first implemented without using Spark framework, achieving an accuracy of 68% and 74%, respectively. These models were then trained on the Spark machine learning framework, achieving better results with an accuracy of 72% for linear regression and 81% for Random Forest models. The batch processing approach was used without taking advantage of Spark's real-time processing capabilities.

In another field, sentiment analysis is the most common trend in today's world, where a lot of work has been done in this sector. Social media platforms are a vital source of information in this case, and Twitter receives millions of tweets every day. Different industries attempt to use this massive textual data to extract people's opinions towards their products. Hadoop framework was used to analyze a large number of tweets

that express users' opinions, classify them, and allocate polarity to each tweet, whether the user expresses positive or negative opinions[15]. The average accuracy in classifying opinion as positive, negative, or neutral was 72.27, attempting to achieve an acceptable response time by reducing disk access operations.

B. IoT data

The evolution of electronics, networks, and communication technology has paved the way for the integration of IoT in various fields such as life, service, industry, and healthcare. To effectively manage IoT-generated data, a framework has been proposed that leverages distributed data processing[16]. This framework comprises multiple layers, with a primary focus on addressing challenges related to smart cities and decision-making capabilities. The first layer is responsible for communication and data generation, while the second layer focuses on collecting and storing data in a distributed environment. The third layer employs big data techniques like MapReduce for processing the stored data. MapReduce is a parallel processing model that offers high scalability and is widely used in the Hadoop system.

The analysis layer is the last layer and provides a means for people and devices to interact directly to make decisions, predictions, create reports, and recommendations. This study has mapped out a roadmap for researchers in the field of big data processing resulting from IoT. However, it has not been applied to real data, and thus the accuracy of the results cannot be verified.

The importance of conserving energy cannot be overstated, and as such, a proposed solution for smart homes is the implementation of an Energy Management System (EMS)[17]. Under this system, each household device is linked to a data acquisition unit, with every device functioning as an IoT object and assigned a unique IP address. This results in a vast wireless network of devices. The data acquisition unit retrieves energy consumption data from every smart device in each household and transmits the data to a central server for analysis and further processing. This data accumulates as big data in the server, which is then analyzed using Business Intelligence (BI) software and big data analytics to optimize energy consumption and meet consumer demand. Given that air conditioning accounts for 60% of electricity consumption in Arab Gulf states, HVAC units (heating, ventilation, and air conditioning) were utilized as a case study to validate the effectiveness of the proposed system. A preliminary model was developed and tested in the laboratory to simulate HVAC systems in small residential areas. Homeowners are able to monitor their daily, monthly, and yearly energy consumption through graphical charts, while officials in the city's electricity sector can review home consumption in specific areas using geographic charts. It was observed that an increasing number of customers negatively impacts data transfer and resource consumption, ultimately increasing access time. The study primarily focused on the infrastructure, transport protocols, and display methods.

In the field of smart tourism and sustainable cultural heritage, the TreSight system [18] has been introduced in Trento, Italy, combining the IoT and big data analytics. This system provides recommendations for tourists by utilizing data from OpenData Trentino, including points of interest, climate, recommended typical restaurants, and more. It has been further enhanced with additional data collected through sensors dispersed throughout the city, including a wearable wristband, to provide more detailed and precise information related to weather, crowd levels, and tracking tourist activities. Hadoop and other tools were used for processing both static and dynamic data to provide advanced analysis and knowledge extraction. The TreSight system is an effective system currently in use in the city of Trento.

In the field of environment, with the spread of pollutants and the deterioration of air quality, the need to take action to reduce pollution and the emission of toxic gases has become imperative. Thus, it was necessary

to use IoT sensors to measure the level of each pollutant in the air. The possibility of integrating big data and IoT concepts in predicting air pollution caused by harmful gases such as SO₂, NO₂, and others was studied. An alert is raised when a certain threshold (critical level) is reached. Apache Spark and its integrated library, Spark MLlib, were used to build the prediction model [19]. Data stored previously were used without delving into the importance of real-time data processing, with a root-mean-square error (RMSE) of 0.13, which is considered good as long as it is less than 0.3.

From a human perspective, the IoT is being used in conjunction with big data analytics to monitor patients in remote areas. Vital patient data is collected using IoT sensors and transmitted in real-time. The information is immediately analyzed using machine learning techniques in Apache Mahout to determine the severity of the patient's condition within an acceptable response time[20]. Analysis results are then transmitted to nearby primary healthcare physicians, as well as the responsible physician and the patient's family, enabling prompt transfer of critically ill patients to the hospital or immediate response by the physician for first aid and treatment in less severe cases. Ongoing studies continue to explore ways to reduce response time and achieve greater accuracy in matters related to human life.

In industry, IoT and big data analytics go hand in hand to play a significant role in industrial sectors. Analyzing the performance of the two technologies is essential for achieving sustainability in industrial production. Updating industries with machines equipped with embedded sensors has led to a rapid increase in data generation and the need for processing. Hadoop has been proposed as a system for dealing with vast amounts of data, along with machine learning tools for predicting faults and other issues[21]. This has resulted in energy savings, good response times, and minimal delays, while increasing overall production and reducing waste output.

VI. METHODOLOGY

In this section, a comparison is conducted between the performance of Hadoop and Spark for big data analytics, focusing on Spark's privacy capabilities in handling streaming data obtained from the IoT. The performance comparison is executed on a ready-made CSV file of data sourced from an American government organization, the U.S. Environmental Protection Agency (EPA), which was collected using IoT sensors distributed throughout various regions in the United States. These sensors measure the levels of gases in the air, particularly toxic gases such as NO₃, SO₂, CO, and O, every day from the year 2000 to 2016.

This dataset comprises more than 1.7 million records with 29 columns containing information such as the date, location number, and detailed information about the sensor location (state, county, and area). It also includes columns containing the recorded values for each gas (NO₃, SO₂, CO, and O), as well as their quality indicators.

Prior to conducting the experiment, the data underwent pre-processing to identify and clean it. The open-source Pandas library, which was built using the powerful, flexible, and user-friendly Python language, was used to analyze and process the data. Pandas provide numerous features that facilitate data file analysis, exploration, and modification. The file was read, its contents reviewed, and variable patterns identified. The necessary columns, such as date and gas quality index, were retained, and their patterns were converted from string values to numeric values. Additionally, a large number of records were found to contain null values in the SO₂ and CO columns, requiring processing through replacement with the most common value within the column. A new column was also created that represents the overall air quality index (AQI) based on the

quality indicators for all gases. The experiment was performed using a Windows 11 laptop computer, and Table 1 in the specifications shows the computer's details.

Table 1. Device specifications

Num	Device	Specifications
1	processor	Intel Cori7 -11370H @3.3GHz
2	RAM	8.00 GB
3	Hard	512 SDD

The Spark framework version 3.1.1 was used along with Jupyter command-line interface, and the Hadoop framework version 3.1.1 was used with the Windows 10 operating system command prompt. Python language was used in both frameworks to read the dataset and find the number of records in the dataset. Then, statistical information, including the maximum value NO2_max, the minimum value NO2_min, and the standard deviation NO2_sd for the NO2 gas variable, were calculated. The time taken for each framework was recorded separately. The experiment was conducted on the Hadoop framework using the map-reduce processing method, which is the adopted processing method for this framework.

We conducted performance testing for both frameworks in machine learning operations, particularly in machine learning algorithms. We used the built-in sparkML library with the Spark framework, while we needed external libraries to accomplish the same work on the Hadoop framework. We built two classification models: the first using the Random Forest Classifier algorithm, and the second using the Decision Tree Classifier algorithm. We trained the models using the columns we selected, which included all the gas quality index columns and the overall air quality index column. We later tested them using all the gas quality index columns to give us the overall air quality index.

It's worth noting that we simulated the process of receiving IoT data in the Spark system using the Spark Streaming component, which allows for real-time data reception and processing. This was not possible in the Hadoop system due to a lack of components that support real-time streaming operations. To achieve the data streaming topic as if it were coming from real sensors, we used Netcat, a tool that reads and writes data over network connections using the UDP or TCP protocol. We opened a port with an identifier of 9999 and wrote the data to it, with Spark listening on the port and receiving the data to display it as an interactive graph within a web browser. We used Plotly, an open-source interactive library that helps to draw various types of charts, to draw the interactive chart. It includes the hover tool capabilities, which allow us to detect any outliers or anomalies in a large number of data points. We also used Dash, a Python framework created by Plotly for building interactive web applications, developed using Flask web framework and Plotly.js and React.js libraries. Dash is open source and displays its results in a tab within a web browser. Thus, we were able to display the chart within a tab in the browser. By doing so, we were able to display the graphical chart in a tab within the browser. The following are the most important results we obtained from analyzing the data and obtaining NO2 levels in the dataset, as well as machine learning using both mentioned frameworks, Hadoop and Spark, and also displaying streaming data in Spark.

VII. RESULTS

This involves reading a dataset and finding the number of records in the dataset, then obtaining statistical information such as the maximum value, minimum value, and standard deviation of the variable that represents the NO₂ gas values. The data will be processed to identify the key statistical measures related to the variable of interest.

A. Execution time

The execution time refers to the time taken to obtain the mentioned statistical information for the analysis of NO₂ levels when using both Hadoop and Spark frameworks from the beginning of data reading until printing the results and obtaining the report.

Table 2 shows the execution time for both frameworks.

Table 2. Execution time

frameworks	Execution time/second
Hadoop (map-reduce)	136.27
Spark	67

Hadoop took twice the amount of time that Spark system took to perform the same task. Specifically, Hadoop took two minutes and a few seconds, whereas Spark took only a minute and a few seconds. The time taken by both frameworks is illustrated in Figure 4, which clearly demonstrates the difference between them.

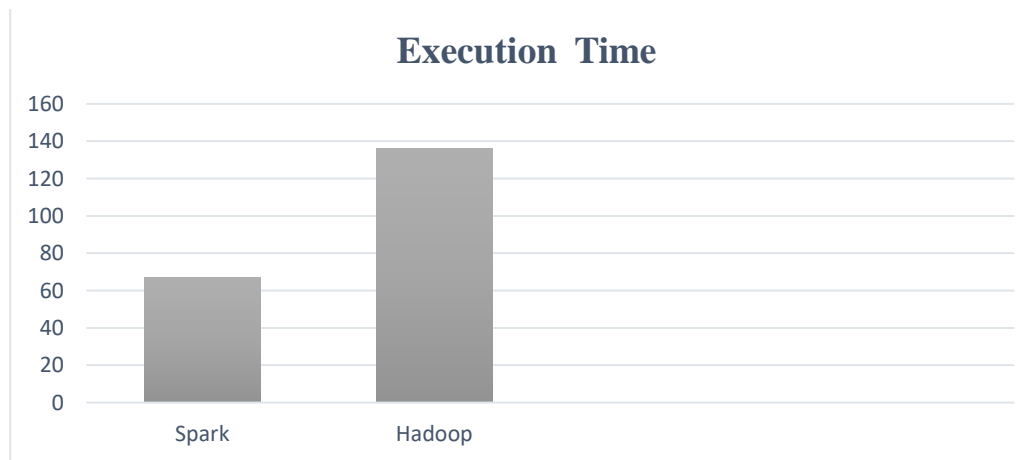


Figure 4. Time taken to perform the task.

1. Resource utilization

It is well-known that the Map-Reduce programming paradigm operates in two stages: mapping, followed by reducing. The intermediate output generated by the mapper is stored on the hard disk, which is then read by the reducer to complete the process[22]. On the other hand, the in-memory processing approach adopted by Spark retains data in memory without the use of the hard disk. This is the main reason for the speed in task performance using Spark. Apache Spark provides a web interface (Spark UI) to monitor the status of the running Spark application (jobs, stages, tasks, etc.) and resource consumption. It is available at [http://\[driver\]:4040](http://[driver]:4040) by default. By monitoring the performance of the experiment using Spark UI, we can see from Table 3 that the task was completed using only memory without the need for the hard disk during execution.

Table 3. Use the hard drive in Spark.

Address	Statuses	RDD Blocks	Storage Memory	Disk Used	Cores
DEDKTOP-H9KELC7:12665	Active	0	88.4 KiB/ 366.3 MiB	0.0B	1

B. Second Experiment (Iterative Processes)

The significance of the framework's capacity to handle multiple tasks, such as machine learning operations, is reflected in numerous instances. It is widely known that these tasks perform optimally when computational memory is employed, as evidenced by the results in Table 4 when using the Random Forest Classifier algorithm.

Table 4. Execution time Random Forest Classifier.

frameworks	Execution time/second
Hadoop	240
Spark	67

Based on the information provided, it appears that Spark had the best performance compared to Hadoop when executing the Decision Tree Classifier algorithm, which consumed less time than its predecessor. The results, as shown in Table 5, demonstrate Spark's superiority as well.

Table 5.. Execution time decision Tree Classifier.

frameworks	Execution time/second
Hadoop	90
Spark	58

Figure 5 illustrates the performance of both frameworks in machine learning and shows the difference between them.

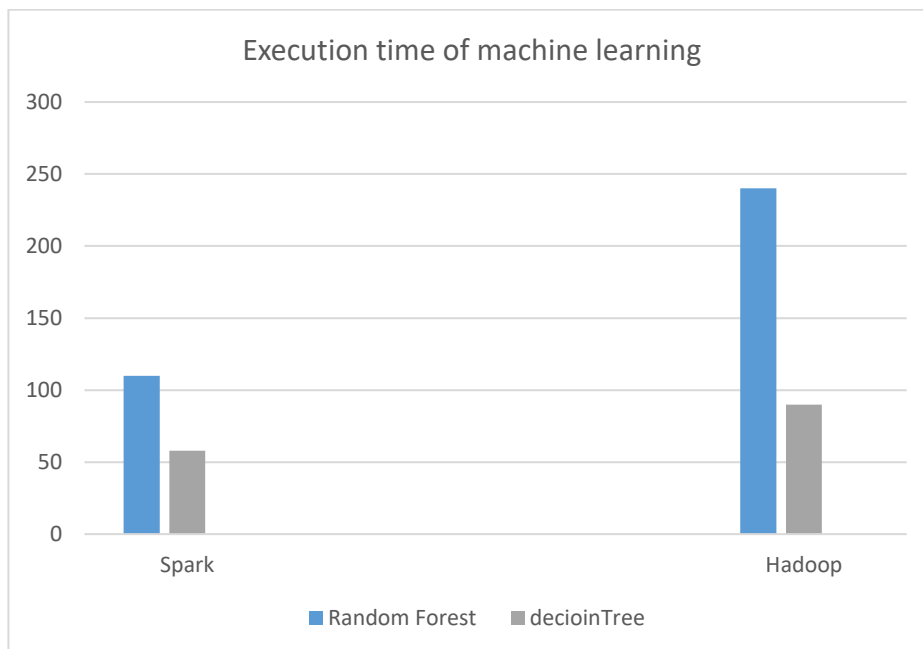


Figure 5. Execution time of machine learning.

C. Third experience (data flow)

Here, Spark's uniqueness in dealing with IoT data is evident, as it is characterized by a constant flow of data that we were able to simulate, analyze, and present graphically in real-time. We created a Live Chart that constantly updates to depict the new value of NO2 gas upon arrival, as shown in Figure 6. The X-axis represents the moment of arrival, while the Y-axis represents the value received. The chart is interactive, allowing us to obtain the recorded value at a specific point by hovering over it with the mouse cursor.

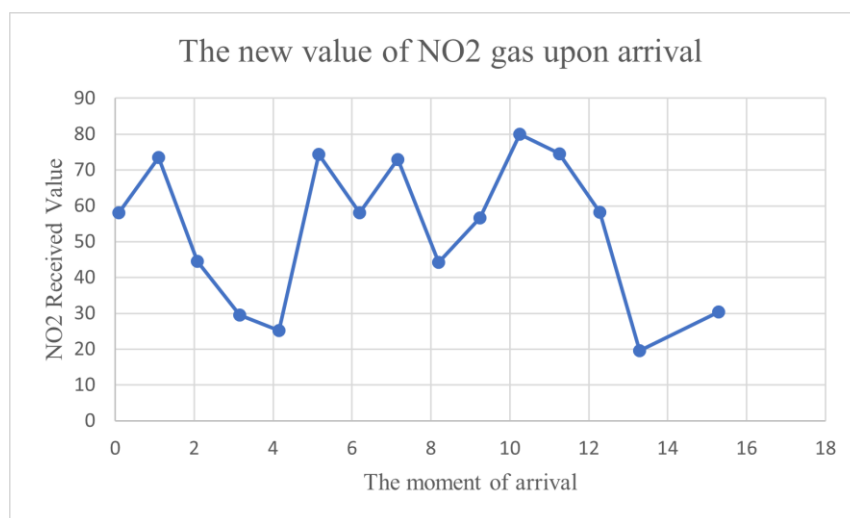


Figure 6. The flowing NO2 gas values.

It has been observed that Spark surpasses Hadoop in terms of speed, completing tasks in almost half the time required by Hadoop. Additionally, Spark operates using memory without resorting to the hard disk. Spark also boasts the integrated machine learning library, MLlib, and the Spark Streaming component, which enables us to receive and process streaming data in real-time.

VIII. DISCUSSION

A. Execution Time

The results presented in Table 2 clearly indicate a substantial difference in execution time between Hadoop and Spark frameworks. Hadoop took 136.27 seconds, while Spark completed the same task in 67 seconds. This stark contrast highlights the efficiency of Spark over Hadoop in processing and analyzing NO2 gas levels. The graphical representation in Figure 4 visually reinforces this point. The efficiency gain observed in Spark can be attributed to its in-memory processing approach, as opposed to Hadoop's reliance on intermediate data storage on the hard disk. This difference in resource utilization is a key factor contributing to Spark's faster execution time.

B. Second Experiment (Iterative Processes)

In the second experiment, where iterative processes, specifically machine learning operations, were employed, Spark once again demonstrated superior performance over Hadoop. Table 4 reveals that Spark outperformed Hadoop significantly in executing the Random Forest Classifier algorithm, taking only 67 seconds compared to Hadoop's 240 seconds. The advantage persisted in the Decision Tree Classifier algorithm, as shown in Table 5, with Spark completing the task in 58 seconds compared to Hadoop's 90 seconds. The graphical representation in Figure 5 underscores Spark's dominance in handling machine learning tasks efficiently. The ability to handle iterative processes with such speed is crucial for real-time applications and scenarios that require quick decision-making.

C. Third Experience (Data Flow)

The third experiment focused on Spark's capability to handle a constant flow of IoT data. As illustrated in Figure 6, Spark demonstrated its prowess in processing and presenting real-time data, outperforming Hadoop in terms of speed and responsiveness. The Live Chart depicted the NO2 gas values arriving in real-time, showcasing Spark's ability to analyze and visualize data as it flows. This experiment highlights Spark's

strengths in dealing with streaming data, a crucial aspect in IoT applications. The observed speed advantage, combined with Spark's ability to operate solely in memory without resorting to the hard disk, positions it as a robust solution for real-time data processing.

In conclusion, the experiments clearly demonstrate that Spark, with its in-memory processing, excels over Hadoop in terms of execution time, resource utilization, and handling iterative processes. The ability to efficiently process streaming data further solidifies Spark as a powerful framework for applications requiring speed, responsiveness, and real-time analytics.

IX. CONCLUSION

In conclusion, this paper conducted a comprehensive theoretical comparison between Apache Spark and Apache Hadoop to evaluate their suitability for analyzing IoT data. The comparison highlighted the advantages of Spark in terms of speed, ease of use, real-time data processing, machine learning capabilities, and cost efficiency. Based on these findings, it can be concluded that Spark is the more appropriate framework for analyzing large-scale streaming IoT data, and it has the potential to enable the development of architectures that can analyze real-time data effectively. As IoT data continues to grow in volume and complexity, the importance of using appropriate big data frameworks for IoT data analysis cannot be overstated. Therefore, the findings of this study have significant implications for researchers, practitioners, and decision-makers who are involved in IoT data analysis.

X. FUTURE WORKS

It is necessary to continue working on the Spark framework and present the graphs in a way that displays the necessary statistical values in real-time, allowing for possible critical situations to be addressed.

XI. REFERENCES

- [1] B. Vuelta, "How Much Data Is Created Every Day? [27 Staggering Stats]," *Seed Sci.*, 2021.
- [2] M. A. Makarem and M. A. Razaz, "Role of Big Data Analytics in the Internet of Things: A Comprehensive Survey," in *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, IEEE, Oct. 2023, pp. 1–15. doi: 10.1109/eSmarTA59349.2023.10293480.
- [3] R. Q. Shadad, M. A. Razaz, G. S. Ali, A. A. Mhammed, H. A. Qaid, and M. Abdulatef, "Efficient and Reliable Management of 5G Network Slicing based on Deep Learning," *2022 2nd Int. Conf. Emerg. Smart Technol. Appl. eSmarTA 2022*, 2022, doi: 10.1109/eSmarTA56775.2022.9935393.
- [4] T. Kirste, "Smart environments," *True Visions Emerg. Ambient Intell.*, pp. 321–337, 2006, doi: 10.1007/978-3-540-28974-6_17.
- [5] Oracle IoT, "What Is the Internet of Things (IoT)," *What Is IoT?*, 2020.
- [6] ITSO, "What is smart ticketing?," *Itso*.
- [7] E. Ahmed *et al.*, "The role of big data analytics in Internet of Things," *Comput. Networks*, vol. 129, pp. 459–471, 2017, doi: 10.1016/j.comnet.2017.06.013.
- [8] Priyadarshini, "What It Is and Why It Matters," *Simplilearn*, vol. 9169, pp. 1–10, 2017.
- [9] IBM Cloud Education, "Hadoop vs. Spark: What's the Difference," 2021.
- [10] PhoenixNAP, "Hadoop vs Spark – Detailed Comparison," *PhoenixNAP*, 2020.
- [11] Goran Jevtic, "Hadoop vs Spark: Detailed Comparison of Big Data Frameworks," 2020.
- [12] M. J. Awan, M. S. M. Rahim, H. Nobanee, A. Yasin, O. I. Khalaf, and U. Ishfaq, "A big data approach to black Friday sales," *Intell. Autom. Soft Comput.*, vol. 27, no. 3, pp. 785–797, 2021, doi: 10.32604/iasc.2021.014216.
- [13] A. Wilson, "What's the difference?," *Home Healthc. Nurse*, vol. 12, no. 5, 1994, doi: 10.1097/00004045-199409000-

00001.

- [14] "Random Forest Algorithms - Comprehensive Guide With Examples."
- [15] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and Vader sentiment," *Lect. Notes Eng. Comput. Sci.*, vol. 2239, pp. 12–16, 2019.
- [16] K. K. Mohbey, "An efficient framework for smart city using big data technologies and internet of things," *Adv. Intell. Syst. Comput.*, vol. 714, pp. 319–328, 2019, doi: 10.1007/978-981-13-0224-4_29.
- [17] A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using IoT and big data analytics approach," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 426–434, 2017, doi: 10.1109/TCE.2017.015014.
- [18] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities," *IEEE Access*, vol. 4, pp. 766–773, 2016, doi: 10.1109/ACCESS.2016.2529723.
- [19] S. S. Alaoui, B. Aksasse, and Y. Farhaoui, "Air pollution prediction through internet of things technology and big data analytics," *Int. J. Comput. Intell. Stud.*, vol. 8, no. 3, p. 177, 2019, doi: 10.1504/ijcistudies.2019.102525.
- [20] S. S and J. S. Raj, "Internet of Things and Big Data Analytics for Health Care With Cloud Computing," *J. Inf. Technol. Digit. World*, vol. 01, no. 01, pp. 9–18, 2019, doi: 10.36548/jitdw.2019.1.002.
- [21] M. M. Vopson, "The world's data explained: how much we're producing and where it's all stored," *Conversat.*, no. May, p. 6, 2021.
- [22] "Hadoop - Mapper In MapReduce - GeeksforGeeks."
- [23] Statista, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025," *Statista*, 2020.
- [24] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, and M. Guizani, "Internet-of-things-based smart environments: State of the art, taxonomy, and open research challenges," *IEEE Wirel. Commun.*, vol. 23, no. 5, pp. 10–16, Oct. 2016, doi: 10.1109/MWC.2016.7721736.

Anomaly Detection Techniques in Communication and Network Systems

Lesia Mochurad, Ivan Dubravskyi

*Department of Artificial Intelligence
Lviv Polytechnic National University
Lviv, Ukraine*

Abstract—Anomalies in the realm of communication and network systems pose significant security and reliability challenges. This work explores anomaly detection techniques in communication and network systems. Various methods and approaches aimed at identifying unusual and potentially hazardous events in network communications are investigated. The paper introduces data analysis methods for detecting abnormal changes in communication and network systems. The results obtained demonstrate a minimal occurrence of false alarms across all three datasets, with corresponding rates of 0.0039, 0.0076, and 0.0009, while simultaneously achieving an impressive accuracy of 99.7% across the entirety of the datasets. To strengthen and consolidate the role of artificial intelligence in network security, this work examines how artificial intelligence can be employed in the field of network security. As a result, a new anomaly detection system for communication networks based on ensemble learning has been developed. The findings of this work can be valuable for enhancing the security and efficiency of communication and network systems, particularly in detecting attacks, internal anomalies, and system failures.

Keywords—*Network anomaly, communication systems, communication networks, data analysis, artificial intelligence.*

I. INTRODUCTION

The general concept of a network anomaly refers to an event that deviates from the network's usual behavior [1]. However, because there are no established models for defining what constitutes normal network behavior, developing a highly rigorous anomaly detector is a challenging task. Depending on the inherent complexity of characterizing typical network behavior, the issue of anomaly detection can be divided into two categories: model-based and non-model-based. Model-based anomaly detectors operate on the assumption that a recognized model exists for the regular behavior of specific aspects of the network, and any deviation is treated as an anomaly [2]. In cases where network behavior cannot be described using a model, non-model-based approaches are employed. These non-model-based approaches can be further classified into those tailored to specific implementations and those that impose accuracy constraints on the detector.

In situations where statistical analysis of network anomalies is necessary for network monitoring, having a predefined model is not obligatory. However, a statistical anomaly detector requires access to a substantial volume of data to gather the necessary samples for accurately assessing typical network behavior. Nevertheless, with the increasing speeds of network links, achieving the desired accuracy through sampling rates may become impractical. Monitoring large network links in real-time necessitates the processing of each packet, along with the management of a considerable amount of state information in memory. As a result, resource limitations restrict sampling rates. In such cases, it's more practical to employ anomaly detection methods that can process continuous data streams with minimal memory requirements and limited state information. Thus, resource-constrained online anomaly detection involves specific data queries that are better suited for discrete algorithms capable of handling streaming data. When compared to statistical sampling, stream analysis examines every data element for the most critical information, whereas sampling processes only a small portion of data and retains all the information within them [3].

The goal of our work is to develop data analysis methods for detecting anomalous changes in communication systems and networks. To achieve this goal, we conducted research and analysis of existing anomaly detection methods and also developed new data analysis methods that provide more precise and effective anomaly detection in communication systems and networks.

The research methods include machine learning algorithms used in data analysis and image processing.

The novelty of this research lies in the development of new methods and algorithms for anomaly detection in communication systems and networks. These methods take into account the specific characteristics and capabilities of modern communication systems and networks, making them more effective compared to existing approaches.

The practical value of this work lies in its potential to enhance the reliability and security of communication systems and networks. The developed methods enable the timely detection and mitigation of anomalies, helping to prevent serious disruptions and issues in the operation of networks and communication systems. Additionally, a novel architecture of a combined neural network has been developed, allowing for effective real-time anomaly detection.

II. RELATED WORK

Based on a literature review on the research topic, six articles directly related to the research theme were selected. They are summarized in Table 1 below.

Table 1
Existing Work Review

No	Name	Instrumentation	Problem	Year	Reference
1	An Innovative Approach to Anomaly Detection in Communication Networks Using Multifractal Analysis	MATLAB, R та C++	Detecting anomalies in communication networks using multifractal characteristics analysis.	2020	[4]
2	An adaptive protection system for sensor networks based on analysis of neighboring nodes	Java, Apache Hadoop, Apache Storm	Creating a new method for protecting wireless sensor networks from attacks.	2021	[5]
3	Design and Testing Novel One-Class Classifier Based on Polynomial Interpolation with Application to Networking Security	Python, NumPy та SciPy, Scikit-learn, Wireshark	Classification for solving the problem of anomaly detection in networks using polynomial interpolation.	2022	[6]
4	Anomaly Detection Framework in Fog-to-Things Communication for Industrial Internet of Things	Python, Pandas, Scikit-learn, PCA, t-SNE, k-means, DBSCAN	Ensuring reliability and security in "Fog-to-Things" systems.	2022	[7]
5	Cybersecurity of Industrial Cyber-Physical Systems: A Review	The authors do not use any specific tools or instrumentation in this article. It is a theoretical and	Research on cybersecurity issues in industrial cyber-physical systems.	2022	[8]

review article.

6	Deep Learning-Based Adaptive Compression and Anomaly Detection for Smart B5G Use Cases Operation	Python, TensorFlow, Keras For neural networks LSTM, GRU, and Autoencoder.	Research on adaptive compression and anomaly detection methods based on deep learning for use in 5G networks.	2023	[9]
---	--	--	---	------	-----

In modern communication networks and communication systems, various anomalous changes occur, such as intruder attacks, equipment malfunctions, intentional or unintentional user errors, and more. These anomalies can lead to disruptions in the normal operation of the network and serious communication issues. Therefore, the development of effective data analysis methods for detecting abnormal changes in communication systems and networks is a relevant issue to ensure stability and security in communication networks.

III.METHODOLOGY

The proposed system architecture comprises multiple components for anomaly detection. The initial component in the system involves intrusion detection, with databases situated behind a firewall, and the network data is preprocessed. Post preprocessing, it is essential to identify and replace any null values with alternative values after confirming the presence of missing data in the system. By default, mean values are used, and duplicate entries are eliminated from the dataset. Subsequently, data dimensionality reduction is applied to encoded data to facilitate efficient data management. Feature optimization is then carried out to emphasize the most relevant characteristics for anomaly detection.

The refined data proceeds to the next stage, where a method known as CFS-RF is exclusively employed to select significant features that contribute to the results. To distinguish between legitimate activities and potential attacks, the system utilizes modified hybrid algorithms, namely AdaBoosting and Bagging (HABBAs), as classifiers. This process is broken down into several phases, each of which performs a specific task in a sequential series of stages, with the output of the previous step serving as the input for the subsequent stage. Following an analysis of the collected datasets, including NSL_KDD [10], UNSW_NB2015 [11], and CIC_IDS2017 [12], three primary preprocessing steps, namely filtration, transformation, and normalization, are executed.

Data preprocessing transforms raw data into a suitable format for analysis and machine learning. It consists of three stages:

Filtering Stage

Filtering removes redundant or unnecessary material from the data, simplifying datasets for ease of use and understanding. It classifies the remaining data and redistributes it into categorized groups.

Transformation Stage

Transformation converts categorical feature data into numeric format using algorithms like one-hot encoding. For example, it is used to convert various protocols, such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), into numerical data.

Normalization Stage

To scale numeric values between 0 and 1, the Min-Max function is used to ensure consistent data ranges for analysis.

These preprocessing stages are crucial for converting raw data into a structured format that can be effectively utilized for analysis and machine learning applications.

This algorithm is designed for preprocessing datasets, specifically datasets D1 (NSL KDD), D2 (CIC IDS2017), D3 (UNSW NB2015). Its purpose is to calculate the worth (usefulness) of features in these datasets, likely for feature selection or dimensionality reduction.

Here's a description of the algorithm:

Input: Datasets D1, D2, D3.

Output: X_i , which represents the worth or usefulness of features.

Algorithm Steps:

1. Loop: This algorithm processes each dataset in a loop.
2. Step 1 and Step 2: These steps involve data preprocessing.
 - Filtration: Remove unnecessary or pointless occurrences from the dataset.
 - Data Transformation: Prepare the data for further analysis.
 - Distribution Classification: Organize the data for further processing.
3. Step 3: Normative Adjustment:
 - Using the minimum and maximum values found in the dataset:
 - Determine the highest value (*Max*).
 - Determine the lowest value (*Min*).
 - Calculate the worth (usefulness) of features, using the formula: $X_i = \frac{X_i - Min}{Max - Min}$.
4. Continue these measures until all features in the dataset have been processed.

Return the X_i , which represents the worth or usefulness of features after preprocessing.

This algorithm appears to be a process for feature selection and evaluation using the CFS-RF (Correlation-Based Feature Selection – Random Forest) method. Let's break down the steps:

1. The algorithm begins by applying the CFS-RF algorithm to select a set of "XiBest" features for each dataset. These features are presumably the most relevant features for further analysis.
2. The algorithm proceeds to iterate through each of the selected "XiBest" features.
3. Each "XiBest" feature is assigned an initial weight value of w_i , which is set to 0.
4. A new RF is created by applying a specific equation and generating 10 separate forest models. The weight values for all "XiBest" features remain 0.
5. The algorithm iterates through each "XiBest" feature in the training set, which was produced from the Random Forest created in the previous step.
6. For each "XiBest" feature, the algorithm calculates a value w . If the calculated value σ is greater than 0, the following steps are executed.
7. For each "XiBest" feature, the weight value w is updated.
8. The algorithm concludes by performing measurement computations, including accuracy, detection rate (DR), and false alarm rate (FAR) to assess the performance of the selected features and the Random Forest model.

In order to maximize the effectiveness of these upgraded classifiers, the revised model combines them to operate in parallel as bagging and employs a weighted average voting approach. By adjusting the weights, it transforms the Random Forest (RF) to work sequentially as AdaBoosting. The aggregation technique is then applied to these updated classifiers, using weighted averages to achieve optimal results in terms of variance and bias. This modified algorithm version leads to enhanced outcomes with reduced error rates.

IV.RESULTS

As a result, a system has been developed that is more resilient to new threats and clearly distinguishes between normal and anomalous activities. However, the research has certain limitations: when using attacks from the dataset, the system performs well but does not consider additional attacks launched from external networks (if any).

The number of selected features, False Alarm Rate (FAR), Detection Rate (DR), and the accuracy of the updated method were analyzed and compared with existing research (see Table 2). During testing, the detection accuracy was 90%, while during training, it was 99%. This led to a higher DR value with a lower FAR value compared to the one-step technique.

Furthermore, the system consistently achieved the highest levels of accuracy, DR, and FAR compared to previous research. The detection accuracy during training was 99%, and during testing, it was 90%. This resulted in a higher FAR value compared to DR, which was lower than in the one-

step technique. Additionally, compared to past research, the system consistently reached the highest levels of accuracy, DR, and FAR.

Table 2

Comparison of the system's results with the results of other research

Systems	Dataset	Number of FS	Accuracy, %	FAR, %	DR, %
[13]	NSL_KDD	N/A	81.00	N/A	82.00
[14]	NSL_KDD	16	84.00	4.85	83.01
	NSL_KDD	42		5.12	96.30
[15]	UNSW_NB2015	13	95.20	N/A	96.65
	CIC_IDS2017	8		N/A	98.94
[16]	CIC_IDS2017	N/A	99.00	N/A	N/A
	NSL_KDD		87.00	0.004	99.02
[17]	UNSW_NB2015	30	85.00	N/A	N/A
	NSL_KDD,	30	99.40	0.004	99.90
[18]	UNSW_NB2015	35	99.80	0.008	99.60
	CIC_IDS2017	40	99.70	0.0012	99.40
	NSL_KDD,	30	99.54	0.0038	99.90
Our System	UNSW_NB2015	35	99.93	0.0076	99.73
	CIC_IDS2017	40	99.84	0.0009	99.55

V.DISCUSSION

The hybrid CFS-RF method reduces dimensionality and eliminates unnecessary attributes from the dataset. At this stage, the analysis and distribution of datasets using the proposed CFS-RF method result in 30 features for NSL_KDD, 35 features for UNSW_NB2015, and 40 features for CIC_IDS2017.

Table 3 demonstrates performance measurements and execution times in three states (i.e., all 30, 35, and 40 features). They show that the selected 30, 35, and 40 features are optimal for assessing system measurements (i.e., accuracy, F-measure, and DR detection speed – 99%). Additionally, they are the best features for reducing execution time to 0.539, 0.839, and 0.931 seconds.

Table 3

Performance result CFS_RF

Dataset (features)	Accuracy	F-measure	DR detection rate	Error rate FAR	Execution time
NSL_KDD (30)	0.995	0.995	0.998	0.004	0.539
UNSW_NB2015 (35)	0.99	0.992	0.99	0.008	0.839
CIC_IDS2017(40)	0.998	0.99	0.99	0.001	0.931

Testing on three datasets allows for a comparison of the proposed CFS-RF with several common feature selection methods, such as Information Gain (IG), Information Gain Ratio (IGR), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Neural Network (NN), and Auto Encoder (AE) (see Figure 1).

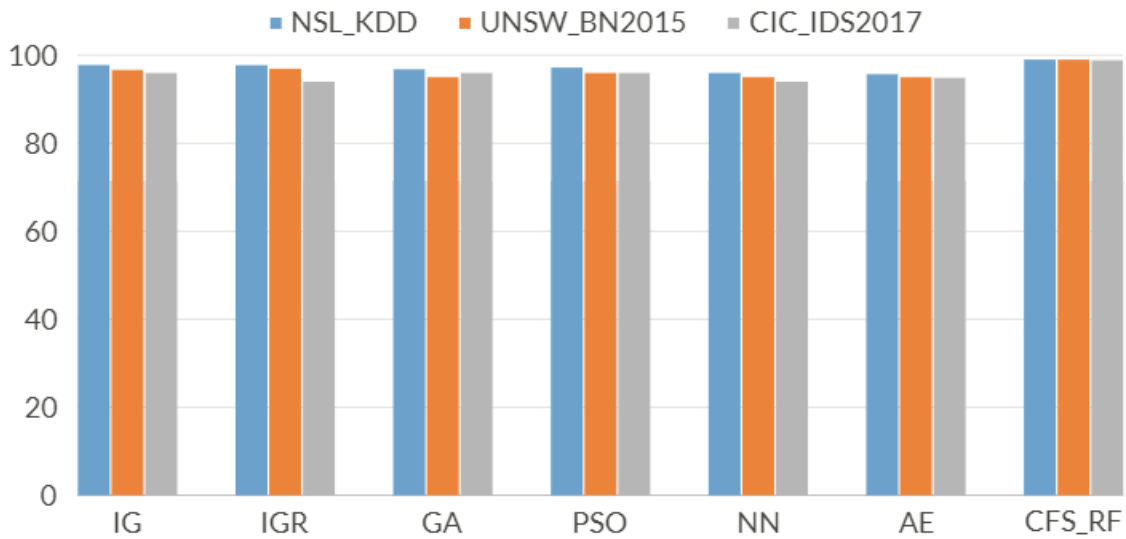


Fig. 1. Comparison of CFS_RF accuracy with other feature selection methods

Table 4 displays the accuracy of each attack in the dataset measured by the F-measure. It also highlights the best results for all classes when applying the proposed system, reaching 100% for XXS and Bot. This indicates that the number of features is ideal and beneficial for identifying all forms of attacks.

Table 4

Attack accuracy and F-measure when using the proposed system

id	Types of attacks	Accuracy, %	F-measure, %
0	DDoS	98	99
1	Port Scan	100	99
2	Brute Force	98.4	100
3	Dos Hulk	99	99
4	DoS GoldenEye	98.9	99
5	Bot	100	100
6	DoS slow loris	99	99
7	PTP Patator	98	99
8	SSH-Patator	99	99
9	XSS	100	100
10	Benign	99	99

VI.FUTURE RESEARCH

For further research, it is possible to thoroughly study and apply more machine learning and deep learning methods to address various security challenges in different domains, such as the Internet of Things, smart cities, API call-based approaches, mobile communication networks, and smart power grids.

VII.CONCLUSION

Contemporary Intrusion Detection Systems continue to face challenges in their effectiveness,

mainly due to vulnerabilities in emerging wireless paradigms. Despite the integration of various machine learning techniques to enhance their performance, these systems remain susceptible to threats. In this research, we leveraged the datasets, which encompasses a broader spectrum of attacks, including DDoS, DoS, brute force, XSS, SQL injection, botnets.

We developed a novel IDS technique that addresses imbalanced feature selection and employs hybrid strategies based on high-dimensional traffic with reduced detection rates. This involved using samples from datasets like NSL KDD, UNSW NB 2015, CIC_IDS2017 to extract 30, 35 and 40 optimal features, respectively.

Our results show a low false alarm rate for all datasets, with values of 0.0039, 0.0076, 0.0009, respectively, while maintaining an impressive accuracy of 99.7% across all datasets. The results comparison table provides additional parametric insights. Notably, the number of selected records inversely correlates with the difficulty level of categorization in the original KDDCup'99 dataset. This method offers an excellent and dependable approach for swiftly and accurately identifying network breaches.

VIII.ACKNOWLEDGMENT

The authors would like to thank the Armed Forces of Ukraine for providing security to perform this work. This work has become possible only because of the resilience and courage of the Ukrainian Army.

IX.DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

REFERENCES

- [1] J. Zhang, R. Gardner, I. Vukotic, "Anomaly detection in wide area network meshes using two machine learning algorithms," *Future Generation Computer Systems*, vol. 93, 2019, pp. 418-426, doi:10.1016/j.future.2018.07.023.
- [2] M. G. Ciobanu, F. Fasano, *et al.*, "Model Checking for Data Anomaly Detection," *Procedia Computer Science*, vol. 159, 2019, pp. 1277-1286. doi:10.1016/j.procs.2019.09.297.
- [3] G. Margetis, G.Tsagakatakis, S. Stamou, C. Stephanidis, "Integrating Visual and Network Data with Deep Learning for Streaming Video Quality Assessment," *Sensors*, 23(8):3998, 2023, doi:10.3390/s23083998.
- [4] P. Dymora and M. Mazurek, "An Innovative Approach to Anomaly Detection in Communication Networks Using Multifractal Analysis," *Applied Sciences*, 10(9):3277, 2020, doi: 10.3390/app10093277.
- [5] J. Gamec, E. Basan, A. Basan, A. Nekrasov, C. Fidge, and N. Sushkin, "An Adaptive Protection System for Sensor Networks Based on Analysis of Neighboring Nodes," *Sensors*, 21(18):6116, 2021, doi: 10.3390/s21186116.
- [6] P. Dini *et al.*, "Design and Testing Novel One-Class Classifier Based on Polynomial Interpolation With Application to Networking Security," *IEEE Access*, vol. 10, pp. 67910–67924, 2022, doi:10.1109/ACCESS.2022.3186026.
- [7] T. Alatawi and A. Aljuhani, "Anomaly Detection Framework in Fog-to-Things Communication for Industrial Internet of Things," *Computers, Materials & Continua*, 73(1):1067–1086, 2022, doi: 10.32604/cmc.2022.029283.
- [8] H. Kayan, M. Nunes, O. Rana, P. Burnap and Ch. Perera, "Cybersecurity of Industrial Cyber-Physical Systems: A Review," *ACM Comput. Surv.*, 54, 35 pages, 2022, doi:10.1145/3510410.
- [9] A. El Sayed, M. Ruiz, H. Harb, and L. Velasco, "Deep Learning-Based Adaptive Compression and Anomaly Detection for Smart B5G Use Cases Operation. *Sensors*," 23(2):1043, 2023, doi: 10.3390/s23021043.
- [10] UNSW-NB15 a network intrusion Dataset. Available online: <https://cloudstor.aarnet.edu.au/plus/index.php/s/2DhnLGDdEECo4ys?path=%2FUNSW-NB15%20-%20CSV%20Files> (accessed on 2023).
- [11] NSL-KDD Network Security Dataset. Available online: <https://www.kaggle.com/datasets/hassan06/nslkdd> (accessed on 2023).
- [12] CIC_IDS2017 labeled network flows Dataset. Available online: <https://www.kaggle.com/datasets/cicdataset/cicids2017/code> (accessed on 2023).
- [13] S. Gadal, R. Mokhtar, M. Abdelhaq, R. Alsaqour, E.S. Ali, R. Saeed, "Machine Learning-Based Anomaly Detection Using K-Mean Array and Sequential Minimal Optimization," *Electronics*, 11(14):2158, 2022, doi:10.3390/electronics11142158.
- [14] Johnson J.M., Yadav A., "Fault Location Estimation in HVDC Transmission Line Using ANN," In Proceedings of the First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1(Smart Innovation, Systems and Technologies), Ahmedabad, India, 28–29 November 2015; pp. 205–

211.

- [15] J.M Johnson, A. Yadav, "Fault Location Estimation in HVDC Transmission Line Using ANN," In Proceedings of the First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1(Smart Innovation, Systems and Technologies), Ahmedabad, India, 28–29 November 2015; pp. 205–211.
- [16] H.H. Pajouh, G. Dastghaibyfar, S. Hashemi, "Two-tier network anomaly detection model: A machine learning approach," *J. Intell. Inf. Syst.*, 48:61–74, 2015.
- [17] N.K. Kanakarajan, K. Muniasamy, "Improving the Accuracy of Intrusion Detection using Gar-Forest with Feature Selection," In Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015; Springer: Cham, Switzerland, pp. 539–547, 2016.
- [18] O.O. Khalifa, M.H. Wajdi, R.A Saeed, A.H.A. Hashim; M.Z. Ahmed; E.S. Ali, "Vehicle Detection for Vision-Based Intelligent Transportation Systems Using Convolutional Neural Network Algorithm," *J. Adv. Transp.*, 2022, 9189600, 2022.

Формування цільової функції визначення оптимальної зарядки літій-іонних акумуляторів

Сергій Буряк, Оксана Гололобова

*Кафедра автоматики та телекомунікацій
Український державний університет науки і технологій
Дніпро, Україна*

Анотація—Одним із найважливіших експлуатаційних питань використання літій-іонних акумуляторів є визначення оптимального рівня струму зарядки. Від його значення залежить швидкість відновлення заряду, а отже і час повернення акумуляторів до виконання покладених на них завдань. Скорочення тривалості зарядки можна досягти шляхом збільшення струму заряду. Проте це призводить до збільшення руйнівних процесів усередині акумуляторів та може стати причиною їх передчасного виходу з ладу. Метою даного дослідження є встановлення такого значення струму, при якому зарядка акумуляторів буде тривати якнайменше за умови відсутності їх перевантаження в процесі, для збереження закладеного продуктивного ресурсу.

Ключові слова—*Акумулятор, Режими роботи, Струм зарядки, Перевантаження, Оптимальні умови, Строк експлуатації.*

I. ВСТУП

Літій-іонні акумулятори стали однією з найбільш затребуваних акумуляторних технологій завдяки їх високій щільності енергії та потужності, тривалому терміну служби та низькій швидкості саморозряду [4]. Проте вони мають свої особливості. Однією з найважливіших є режим зарядки, для якого характерно обмеження в спроможності відновлювати заряд за струмом і напругою. Надмірні значення цих складових можуть пошкодити, або вивести з ладу накопичувач. З іншого боку дане обмеження призводить до збільшення тривалості заряду акумулятора, що стримує його подальше поширення та популяризацію. Метою дослідження є встановлення оптимальних умов відновлення заряду літій-іонних акумуляторів з мінімізацією можливості руйнувань, викликаних надмірними значеннями показників електричного струму.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Досягнення оптимальної зарядки акумулятора з урахуванням різних обмежень для безпечної, ефективної та надійної роботи є складною задачею. Вона розглядається як складова функція, що враховує три цілі, а саме: час зарядки, втрати енергії та підвищення температури як усередині, так і на поверхні [7]. Як і температура акумулятора на поверхні, так і її внутрішня температура можуть перевищувати допустимий рівень при зарядці великим струмом, що прискорює процес старіння акумулятора і навіть може викликати вибух чи пожежу в надзвичайних ситуаціях [8].

Найпопулярнішою стратегією зарядки літій-іонних акумуляторів стала стратегія – постійний струм – постійна напруга. У цій стратегії зарядка відбувається в два етапи. Спочатку на акумулятор подається постійний струм, при цьому напруга на його клеммах збільшується доти, доки не досягне максимального безпечного порогу. Потім акумулятор починає заряджатися від постійної напруги, доки ємність акумулятора не досягне цільового значення. Повний заряд відбувається, коли акумулятор досягає порогової напруги та струм падає до 3-5% від номінального [2].

Акумулятор також вважається повністю зарядженим, якщо значення струму стабілізується і вже не може зменшуватись далі. Причиною такої ситуації може бути підвищений саморозряд. Літій-іонний акумулятор, що отримав повністю насичений заряд, зберігатиме підвищену напругу довше, ніж акумулятор, що не отримав насичення. Додавання повного насичення при заданій напрузі збільшує ємність акумулятора всього приблизно на 10%, але й підвищує навантаження через високу напругу, а відтак пришвидшує руйнування складових елементів [1].

Хоча дана стратегія проста у застосуванні, відкритою проблемою є вибір відповідного зарядного струму на її першому етапі. З іншого боку, низький зарядний струм може збільшити час зарядки акумулятора, вплинути на зручність його використання та обмежити сфери застосування. Тому дуже важливо розробити найкращу стратегію на основі даного методу, щоб покращити загальну продуктивність зарядки та гарантувати безпеку експлуатації акумулятора.

III. МЕТОДИ

Зменшення часу зарядки збільшує зарядний струм, що підвищує температуру акумулятора. Відомо, що ця динаміка є ключовим чинником прискорення старіння акумуляторів. Рекомендована швидкість зарядки акумулятора становить від $0,5C$ до $1C$, де C – ємність акумулятора. Повний час зарядки складає близько 2-3 годин. Високошвидкісна зарядка, вище за номінальну швидкість заряду C , негативно впливає на продуктивність акумулятора, прискорюючи знос [5]. Тому виробники рекомендують заряджати струмом не вище $0,8C$, щоб продовжити термін служби акумулятора. Незважаючи на це, більшість силових елементів можуть витримувати більш високу швидкість зарядки з невеликим перевантаженням. При цьому її ефективність практично не втрачається і становить близько 99%, а акумулятор залишається холодним [1].

Літій-іонний акумулятор не може бути перезарядженим. При повній зарядці струм зарядки має бути вимкнений. Безперервна крапельна зарядка може призвести до утворення металевого літію та поставити під загрозу безпеку його подальшого використання. Якщо пристрої перебувають у стані підзарядки в увімкненому положенні, то струм, що проходить через пристрій, називається паразитним навантаженням і може спотворити цикл зарядки. Потрібно уникати паразитних навантажень під час зарядки, оскільки вони викликають міні-цикли. Рівень навантаження на акумулятор при цьому високий, оскільки цикли відбуваються на порозі високої напруги [3].

Акумулятор повинен бути вимкнений під час зарядки. Це дозволяє йому безперешкодно досягати встановленого порога напруги та точки насичення струмом. Паразитне навантаження заважає зарядному пристрою, знижуючи напругу акумулятора та не дозволяючи струму на стадії насичення впасти досить низько за рахунок струму витоку. Акумулятор може бути повністю заряджений, але умови роботи при підключеному зарядному пристрої вимагають продовження зарядки, що призводить до перевантаження.

Недоцільно оцінювати заряд шляхом визначення напруги акумулятора, який знаходиться у процесі зарядки. Найбільш інформативно вимірювати напругу холостого ходу після того, як акумулятор перебував у стані спокою протягом декількох годин. На напругу холостого ходу впливає температура, а також активний літій-іонний матеріал. Особливістю процесу першої зарядки акумулятора є різке зростання напруги. Ємність зрештою відновиться, коли акумулятор буде майже повністю заряджений. Низькі температури або зарядка елемента з високим внутрішнім опором підсилюють цей ефект [6, 8].

Щоб сформулювати оптимізовану зарядку акумулятора, необхідно визначити деякі показники продуктивності. Час зарядки акумулятора є ключовим показником ефективності зарядки, і бажано, щоб він був якомога коротшим. Ще одним ключовим показником є втрати енергії акумулятора (споживана потужність) у процесі зарядки. Великі втрати енергії призводять до низької ефективності зарядки акумулятора. Однак час зарядки акумулятора та

втрати енергії є двома суперечливими цілями. Крім того, підвищення температури як всередині акумулятора, так і на його поверхні під час зарядки, також є важливим показником, який необхідно враховувати в процесі зарядки. Слід зазначити, що різниця температур усередині та поверхні акумулятора в процесі зарядки може бути дуже значною. Надмірна температура, особливо внутрішня, призводить до значного пошкодження продуктивності та терміну служби акумулятора і навіть може призвести до серйозних проблем з безпекою. Отже, час зарядки акумулятора, втрати енергії та підвищення температури (як усередині, так і на поверхні) повинні враховуватися в цільовій функції для оптимізації процесу зарядки.

Високий струм може призвести до великих втрат енергії, а температура може перевищити допустимі рівні, особливо у споживачів з високою потужністю. Слід згадати, що деякі літій-іонні акумулятори все ж таки можуть зазнавати підвищення температури приблизно на 5°C (9°F) при досягненні повної зарядки. Таку ситуацію можна вважати нормою, оскільки дане явище може бути пов'язано із захисною схемою та/або підвищеним внутрішнім опором. Припинення використання акумулятора або зарядного пристрою потрібно тоді, коли його температура підніметься більш ніж на 10°C (18°F) при помірній швидкості зарядки [1].

IV. РЕЗУЛЬТАТИ

Збільшення струму зарядки не прискорює стан повного заряду. Хоча акумулятор досягає піку напруги швидше, проте заряд насичення займе відповідно більше часу в наступному кроці. При вищому струмі перша стадія коротша, але насичення на другій стадії займе більше часу. Однак, зарядка великим струмом швидше заповнить акумулятор і його заряд в кінці першої стадії вже може досягти значення біля 75%.

V. ОБГОВОРЕННЯ

Літій-іонний акумулятор не вимагає повної зарядки, як у випадку зі свинцево-кислотним акумулятором, і це навіть небажано. Насправді краще його не заряджати повністю, тому що висока напруга є надмірним навантаженням для нього. Вибір нижчого порога напруги або повне виключення заряду насичення продовжує загальний термін служби акумулятора, але, звісно, скорочує ефективний час роботи. Зарядні пристрої для споживчих товарів розраховані на максимальну потужність та не підлягають регулюванню, а збільшений термін служби вважається для них менш важливим. Проте такий підхід неможливий на промисловому виробництві, у космічній галузі та в усіх інших сферах, де головним параметром функціонування накопичувачів електричної енергії є тривалість загального життєвого циклу, а не максимальна продуктивна здатність одного циклу заряд-розряд.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Концепція оптимізації зарядки літій-іонних акумуляторів надає можливість визначення вагомих умов, які впливають на стан акумуляторів та тривалість відновлення заряду. Вони залежать як від експлуатаційних умов застосування, так і від особливостей побудови, що може бути враховано та розвинуто в подальших вишукуваннях.

VII. ВИСНОВКИ

Таким чином, великий струм зарядки дозволяє досягти більшого значення стану заряду акумулятора на першому етапі зарядки – до моменту досягнення порогового значення напруги, однак не впливає на загальну тривалість зарядки акумулятора до повної зарядки, оскільки в такому випадку збільшується тривалість другого етапу зарядки – до повного насичення. Отже, при визначенні оптимальних умов зарядки акумуляторів слід керуватися такими значеннями струму, при яких не відбувається їх надшвидке старіння, щоб не скоротити закладений термін експлуатації. Це дозволить не збільшувати час зарядки понад необхідного та з іншої сторони не призведе до пошкоджень акумуляторів.

Також слід зазначити, що визначальним чинником для зарядки акумулятора є температура, оскільки занадто високе або низьке її значення може стати причиною його пошкодження.

ЛІТЕРАТУРА

- [1] Q. Lin, J. Wang, R. Xiong, W. Shen, and H. He, "Towards a smarter battery management system: A critical review on optimal charging methods of lithium ion batteries," *Energy*, vol. 183, 2019, pp. 220-234, doi: 10.1016/j.energy.2019.06.128
- [2] K. Liu, K. Li, Z. Yang, C. Zhang, and J. Deng, "An advanced Lithium-ion battery optimal charging strategy based on a coupled thermoelectric model," *Electrochimica Acta*, vol. 225, 2017, pp. 330-344, doi: 10.1016/j.electacta.2016.12.129
- [3] C. Zhang, J. Jiang, Y. Gao, W. Zhang, Q. Liu, and X. Hu, "Charging optimization in lithium-ion batteries based on temperature rise and charge time," *Applied Energy*, vol. 194(C), 2016, pp. 569-577, doi: 10.1016/j.apenergy.2016.10.059
- [4] Z. Chu, X. Feng, M. Ouyang, Z. Wang, L. Lu, J. Li, and X. Han, "Optimal charge current of lithium-ion battery," *Energy Procedia*, vol. 142, pp. 1867-1873, 2017, doi: 10.1016/j.egypro.2017.12.577
- [5] Battery University (2021). BU-409: Charging lithium-ion. Retrieved from: <https://batteryuniversity.com/article/bu-409-charging-lithium-ion>
- [6] D.-R. Kim, J.-W. Kang, T.-H. Eom, J.-M. Kim, J. Lee1, and C.-Y. Won, "Rapid-charging solution for 18650 cylindrical lithium-ion battery packs for forklifts," *Journal of Electrochemical Science and Technology*, vol. 9(3), 2018, pp. 184-194, doi: 10.5229/JECST.2018.9.3.184
- [7] G.-J. Chen, Y.-H. Liu, Y.-S. Cheng, and H.-Y. Pai, "A novel optimal charging algorithm for lithium-ion batteries based on model predictive control," *Energies*, 2021, 14(8), 2238, doi: 10.3390/en14082238
- [8] H. Ruan, B. Sun, J. Jiang, X. Su, X. He, S. Ma, and W. Gao, "Optimal switching temperature for multi-objective heated-charging of lithium-ion batteries at subzero temperatures," *Journal of Power Sources*, vol. 562, 2023, Article 232775, doi: 10.1016/j.jpowsour.2023.232775

Modeling of Wave Propagation in Dispersive Media Using New ADE-TLM Method

El Hadi El Ouardy, Hamid Bezzout, and Hanan El Faylali

Department of physics

Laboratory of Electronic Systems Information Processing Mechanics and Energetics

Ibn Tofail University, B.P: 242

Kenitra, Morocco

Abstract—In this paper, we introduce a novel approach for modeling dispersive media by employing the Transmission Line Matrix (TLM) technique enhanced with the symmetrical condensed node (SCN-TLM) technique. This method effectively simulates electromagnetic wave interactions with Lorentz and Drude media through a scattered-field formulation. Our algorithm leverages polarization current within the media and integrates the auxiliary differential equation (ADE) method to handle scattering during the conventional discretization process. By optimizing the number of operations required for each iteration, we have significantly reduced computational overhead. The effectiveness and accuracy of our approach are demonstrated through the agreement between the calculated reflection coefficient utilizing the New ADE-TLM method and analytical solutions for both Lorentz and Drude media.

Keywords—Auxiliary Differential Equation (ADE), Drude medium, Lorentz medium, Transmission Line Matrix (TLM), Scattered-field.

I. INTRODUCTION

The Transmission Line Matrix (TLM) approach was designed to present a direct comparison between the fundamentals of electromagnetic wave propagation and the rules that regulate electrical currents and voltages within a network of interconnected transmission lines [1, 2]. This network, which demonstrates the discretization of the physical environment studied, is composed of interconnected transmission lines, which serve as fundamental components called nodes [3, 4].

In recent years, we report a many TLM-based approaches for the analysis of a dispersive medium. We find: the use of Constant Recursive Convolution (CRC) [5], the Current Density Recursive Convolution (CDRC) technique [6], the Piecewise Linear Recursive Convolution (PLRC) technique [7], the Piecewise Linear Current Density Recursive Convolution (PLCDRC-TLM) technique [8], the application of the Z-transform approach [9] to characterize the electric properties of the dispersive media, and the other technique includes the JE convolution (JEC) with voltage sources method [10] which is applied to model isotropic plasma.

II. RELATED WORK

The auxiliary differential equation (ADE-TLM) technique is employed to model chiral media [11], Lorentz and Drude media [12, 13], by applying the ADE approximation, and it is additionally used to represent Cole-Cole media [14].

This paper introduces a novel advancement in the ADE-TLM framework, offering an innovative formulation designed for modeling environments characterized by Lorentz and Drude properties. This approach utilizes a unique ADE-TLM algorithm that exploits polarization current density and the fundamental relationships between voltage (V) and polarization current density (J). Built upon the SCN-TLM, it employs 12 principal ports to model free space and incorporates 3 additional ports as voltage sources to characterize the properties of Drude and Lorentz media.

The efficacy of the proposed ADE-TLM formulation is demonstrated through excellent agreement between calculated results and analytical solutions for the reflection coefficient of Lorentz and Drude media. This paper further elaborates on the methodology and presents validation results, contributing to the growing body of knowledge in the field.

III. METHODOLOGY

In this methodology section, we delve into the innovative approach of modeling Lorentz and Drude media through the New Auxiliary Differential Equation-Transmission Line Matrix (ADE-TLM) algorithm. This novel technique, aimed at accurately representing dispersive media interactions, involves a sophisticated formulation incorporating both the Lorentz and Drude media characteristics. The core of our methodology lies in the development of equations that seamlessly integrate the auxiliary differential equation concept with the Transmission Line Matrix method. Through a meticulous process, we have devised a comprehensive algorithm capable of capturing the intricate behaviors of both Lorentz and Drude media, offering a robust foundation for our subsequent analyses and simulations.

1. Formulations and equations

The magnetic and electric fields are separated into incident and scattered terms, resulting in [15]:

$$E_{total} = E_{inc} + E_{scat} \quad \text{and} \quad H_{total} = H_{inc} + H_{scat}$$

In the case of an incident field propagating in free space, Ampere's Law can be expressed as follows:

$$\nabla \times E = -\mu_0 \frac{\partial H}{\partial t} \quad (1)$$

The polarization current density is described by Maxwell equations as follows [16]:

$$\nabla \times H = \varepsilon_0 \varepsilon_\infty \frac{\partial E}{\partial t} + \sum_{p=1}^p J_p \quad (2)$$

Where E is the electric field strength, H represents the magnetic field strength, ε_0 is the permittivity of vacuum, ε_∞ is the relative permittivity at infinite frequency, μ_0 is the magnetic permeability of vacuum, J is the polarization current density and p denotes each Lorentz pole-pair or Drude pole.

1.1 Linear Drude

For the Drude media, the following relation is employed for the polarization current density in Eq. (2) [17, 18]:

$$\frac{\partial^2 J_p}{\partial t^2} + \tau_p \frac{\partial J_p}{\partial t} = \varepsilon_0 \omega_p^2 \frac{\partial E}{\partial t} \quad (3)$$

Where τ_p is the relaxation time for each Drude pole and ω_p represents the resonant frequency.

The updated equation is:

$$J_p^{n+1} = a_d J_p^n + b_d (E^{n+1} + E^n) \quad (4)$$

With the coefficients:

$$a_d = \frac{2 - \tau_p \Delta t}{2 + \tau_p \Delta t} \quad (5a)$$

$$b_d = \frac{\varepsilon_0 \omega_p^2 \Delta t}{2 + \tau_p \Delta t} \quad (5b)$$

We obtain the polarization current:

$$J_p^{n+\frac{1}{2}} = \frac{1}{2} (J_p^{n+1} + J_p^n) \quad (6)$$

We obtain by substituting Eq.(4) in Eq.(6):

$$J_p^{n+\frac{1}{2}} = \frac{1}{2} \left[(1 + a_d) J_p^n + b_d (E^{n+1} + E^n) \right] \quad (7)$$

The temporal discretization at step $n+1/2$ of Eq.(2) using Eq.(7) is :

$$\nabla \wedge H^{n+\frac{1}{2}} = \varepsilon_0 \varepsilon_\infty \left(\frac{E^{n+1} - E^n}{\Delta t} \right) + \frac{1}{2} \left[(1 + a_d) J_p^n + b_d (E^{n+1} + E^n) \right] \quad (8)$$

We obtain from Eq.(8) :

$$E^{n+1} = \left(\frac{2\varepsilon_0 \varepsilon_\infty - \Delta t b_d}{2\varepsilon_0 \varepsilon_\infty + \Delta t b_d} \right) E^n + \left(\frac{2\Delta t}{2\varepsilon_0 \varepsilon_\infty + \Delta t b_d} \right) \left(\nabla \wedge H^{n+\frac{1}{2}} - \frac{1}{2} (1 + a_d) J_p^n \right) \quad (9)$$

We convert the electric field to a voltage utilizing the following equation:

$$E^n = \frac{V^n}{\Delta l} \quad (10)$$

Substituting Eq.(10) in Eq.(9), we obtain:

$$V^{n+1} = \left(\frac{2\varepsilon_0 \varepsilon_\infty - \Delta t b_d}{2\varepsilon_0 \varepsilon_\infty + \Delta t b_d} \right) V^n + \left(\frac{2\Delta t \Delta l}{2\varepsilon_0 \varepsilon_\infty + \Delta t b_d} \right) \left(\nabla \wedge H^{n+\frac{1}{2}} - \frac{1}{2} (1 + a_d) J_p^n \right) \quad (11)$$

We introduce this Eq.(12) in Eq.(11) :

$$\nabla \wedge H^{n+\frac{1}{2}} = \frac{\epsilon_0}{2\Delta t \Delta l} \left[\sum V_i^{n+1} - \sum V_i^n - V_{sx}^n \right] \quad (12)$$

The total voltage is written as follows:

$$\begin{pmatrix} V_x^{n+1} \\ V_y^{n+1} \\ V_z^{n+1} \end{pmatrix} = \begin{pmatrix} \frac{2}{4+Y_{ox}} \\ \frac{2}{4+Y_{oy}} \\ \frac{2}{4+Y_{oz}} \end{pmatrix} \begin{pmatrix} \left[V_1^i + V_2^i + V_9^i + V_{12}^i + \frac{1}{2}V_{sx} \right]^{n+1} \\ \left[V_3^i + V_4^i + V_8^i + V_{11}^i + \frac{1}{2}V_{sy} \right]^{n+1} \\ \left[V_5^i + V_6^i + V_7^i + V_{10}^i + \frac{1}{2}V_{sz} \right]^{n+1} \end{pmatrix} \quad (13)$$

Y_{ox} , Y_{oy} , and Y_{oz} indicate the normalized admittances.

By comparing Eqs. (11) and (13), the normalized admittances Y_{ox} , Y_{oy} , and Y_{oz} , and the voltage sources V_{sx} , V_{sy} , and V_{sz} are represented respectively as follows:

$$Y_{ou} = 4 \left(\frac{2\epsilon_0 \epsilon_\infty + \Delta t b_d}{2\epsilon_0} - 1 \right) \quad (14)$$

$$V_{su}^{n+1} = -V_{su}^n - 4 \left[\frac{\Delta t b_d}{\epsilon_0} V_u^n + \frac{\Delta t \Delta l}{2\epsilon_0} (1+a_d) J_p^n \right] \quad (15)$$

With: $u \in \{x, y, z\}$.

1.2 Linear Lorentz

For Lorentz media, the update equations can be written in the same way as the Drude model. The polarization current density term in Eq. (2) [18]:

$$\omega_p^2 J_p + 2\delta_p \frac{\partial J_p}{\partial t} + \frac{\partial^2 J_p}{\partial t^2} = \epsilon_0 \Delta \epsilon_p \omega_0^2 \frac{\partial E}{\partial t} \quad (16)$$

Where, ϵ_s is the static permittivity, ω_0 is the angular resonant frequency for the p'th Lorentz characteristic, and δ_p is the damping factor.

We obtain by using the finite difference time in Eq. (16) at time step n+1:

$$\omega_p^2 \left(\frac{J_p^{n+1} + J_p^n}{2} \right) + 2\delta_p \left(\frac{J_p^{n+1} - J_p^{n-1}}{2\Delta t} \right) + \left(\frac{J_p^{n+1} - 2J_p^n + J_p^{n-1}}{(\Delta t)^2} \right) = \epsilon_0 \Delta \epsilon_p \omega_p^2 \left(\frac{E^{n+1} - E^{n-1}}{2\Delta t} \right) \quad (17)$$

The updated equation is:

$$J_p^{n+1} = \alpha_p J_p^n + \xi_p J_p^{n-1} + \gamma_p \left(\frac{E^{n+1} - E^{n-1}}{2\Delta t} \right) \quad (18)$$

With the coefficients:

$$\alpha_p = \frac{2 - \omega_L^2(\Delta t)^2}{1 + \delta_L \Delta t} \quad (19a)$$

$$\xi_p = \frac{\delta_L \Delta t - 1}{1 + \delta_L \Delta t} \quad (19b)$$

$$\gamma_p = \frac{\varepsilon_0 \Delta \varepsilon_p \omega_L^2 (\Delta t)^2}{1 + \delta_L \Delta t} \quad (19c)$$

The time discretization of Eq. (2) gives:

$$E^{n+1} = E^n + \frac{\Delta t}{\varepsilon_0} \left(\nabla \wedge H^{n+\frac{1}{2}} - J^{n+\frac{1}{2}} \right) \quad (20)$$

We obtain the polarization current from Eq. (18):

$$J_p^{n+\frac{1}{2}} = \frac{1}{2} (J^{n+1} + J^n) \quad (21)$$

We convert the electric field to a voltage utilizing the following equation:

$$E^n = \frac{V^n}{\Delta l} \quad (22)$$

Where Δl is the space step, and V is the electric voltage.

By substituting Eq. (22) in Eq. (18) and Eq. (20) respectively, we find:

$$J_p^{n+1} = \alpha_p J_p^n + \xi_p J_p^{n-1} + \gamma_p \left(\frac{V^{n+1} - V^{n-1}}{2\Delta l \Delta t} \right) \quad (23)$$

$$V^{n+1} = V^n + \frac{\Delta l \Delta t}{\varepsilon_0} \left(\nabla \wedge H^{n+\frac{1}{2}} - J^{n+\frac{1}{2}} \right) \quad (24)$$

We update Eq. (23), and we find by utilizing the average approximation defined in [19]:

$$V^n = \frac{V^{n+1} + V^{n-1}}{2} \quad (25)$$

We obtain by substituting Eq. (25) into Eq. (23):

$$J_p^{n+1} = \alpha_p J_p^n + \xi_p J_p^{n-1} + \gamma_p \left(\frac{V^n - V^{n-1}}{\Delta l \Delta t} \right) \quad (26)$$

We find by utilizing the SCN-TLM method [20] to Eq. (1):

$$\begin{pmatrix} \nabla \wedge H_x^{n+\frac{1}{2}} \\ \nabla \wedge H_y^{n+\frac{1}{2}} \\ \nabla \wedge H_z^{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon_0}{2\Delta t \Delta l} \left[(V_1^i + V_2^i + V_9^i + V_{12}^i)^{n+1} - (V_1^r + V_2^r + V_9^r + V_{12}^r)^n \right] \\ \frac{\varepsilon_0}{2\Delta t \Delta l} \left[(V_3^i + V_4^i + V_8^i + V_{11}^i)^{n+1} - (V_3^r + V_4^r + V_8^r + V_{11}^r)^n \right] \\ \frac{\varepsilon_0}{2\Delta t \Delta l} \left[(V_5^i + V_6^i + V_7^i + V_{10}^i)^{n+1} - (V_5^r + V_6^r + V_7^r + V_{10}^r)^n \right] \end{pmatrix} \quad (27)$$

Applying the charge conservation's laws:

$$\begin{pmatrix} V_1^r + V_2^r + V_9^r + V_{12}^r \\ V_3^r + V_4^r + V_8^r + V_{11}^r \\ V_5^r + V_6^r + V_7^r + V_{10}^r \end{pmatrix}^n = \begin{pmatrix} V_1^i + V_2^i + V_9^i + V_{12}^i \\ V_3^i + V_4^i + V_8^i + V_{11}^i \\ V_5^i + V_6^i + V_7^i + V_{10}^i \end{pmatrix}^n + \begin{pmatrix} V_{sx} \\ V_{sy} \\ V_{sz} \end{pmatrix} \quad (28)$$

Replacing Eq. (27) in Eq. (24), employing Eq. (28), and we find the total voltage by utilizing the Symmetrical Condensed Node (SCN):

$$\begin{pmatrix} V_x^{n+1} \\ V_y^{n+1} \\ V_z^{n+1} \end{pmatrix} = \begin{pmatrix} \frac{2}{4 + Y_{ox}} \\ \frac{2}{4 + Y_{oy}} \\ \frac{2}{4 + Y_{oz}} \end{pmatrix} \begin{pmatrix} \left[V_1^i + V_2^i + V_9^i + V_{12}^i + \frac{1}{2} V_{sx} \right]^{n+1} \\ \left[V_3^i + V_4^i + V_8^i + V_{11}^i + \frac{1}{2} V_{sy} \right]^{n+1} \\ \left[V_5^i + V_6^i + V_7^i + V_{10}^i + \frac{1}{2} V_{sz} \right]^{n+1} \end{pmatrix} \quad (29)$$

Y_{ox} , Y_{oy} , and Y_{oz} indicate the normalized admittances.

The normalized admittances Y_{ox} , Y_{oy} , and Y_{oz} can be written after comparing between Eq. (24) and Eq.(29) as:

$$Y_{ou} = 4(\varepsilon_\infty - 1) \quad (30)$$

And the voltage sources V_{sx} , V_{sy} , and V_{sz} are:

$$V_{su}^{n+1} = -V_{su}^n - \frac{\Delta l \Delta t}{2\varepsilon_0} \sum_{p=1}^p (J_{pu}^{n+1} + J_{pu}^n) \quad (31)$$

IV. RESULTS

To evaluate the efficiency and validity of the New ADE-TLM algorithm presented in Section 2, which includes voltage sources for the Lorentz and Drude media. We simulate the reflection coefficient of plane wave incident from vacuum onto either a single pole-pair Lorentz half-space or a Drude half-space employing a one-dimensional, the results obtained are compared to the analytical solutions to validate and make comparisons.

The TLM network is divided in $(1,1,1000)\Delta l$, with the time step: $\Delta t = 0.4166$ ps, and the space step: $\Delta l = 250\mu\text{m}$.

We partitioned the one-dimensional computational space into 1000 cells, with 500 cells represented to simulating air and the other 500 cells are occupied by Lorentz and Drude media.

Similar to [21], the following parameters describe the Lorentz media:

$$\varepsilon_s = 3.0, \varepsilon_\infty = 1.5, \omega_0 = 2\pi \times 25\text{GHz}, \text{ and } \delta_L = 0.1\omega_0.$$

While the Drude half space is defined by:

$$\varepsilon_s = 3.0, \varepsilon_\infty = 1.0, \tau_0 = 2 \times 10^{10} \text{ s}$$

Fig. 1 shows waveforms calculated using the New ADE-TLM method. These waveforms show the behavior in the free-space of the scattered electric fields when interacting with both the Lorentz and Drude half-space models. This figure demonstrates the presence of attenuation and dispersion influences.

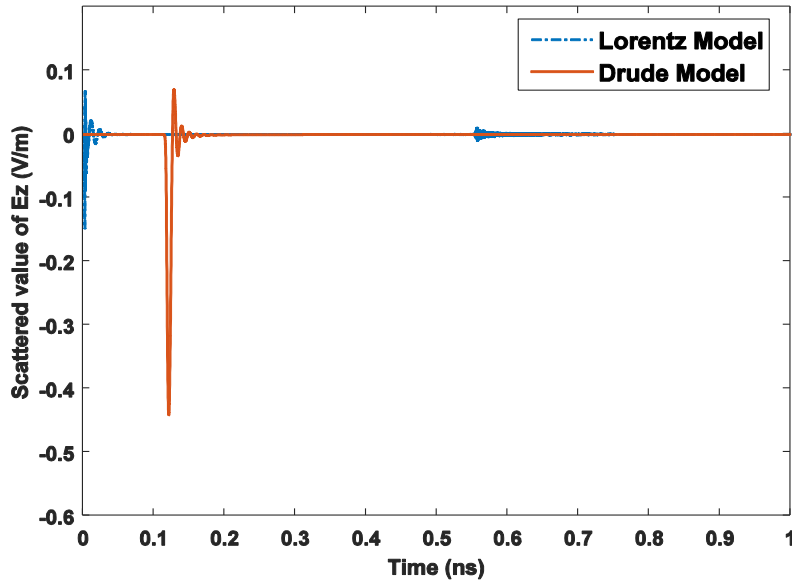


Fig. 1. New ADE-TLM calculated time-waveforms of the scattered electricfield for both the Lorentz and Drude half-space examples

We calculate the analytical solution for the reflection coefficient, utilizing the formulation [22]:

$$|\Gamma(\omega)| = \left| \frac{\sqrt{\varepsilon_0} - \sqrt{\varepsilon^*(\omega)}}{\sqrt{\varepsilon_0} + \sqrt{\varepsilon^*(\omega)}} \right| \quad (24)$$

Where $\varepsilon^*(\omega)$ is the complex permittivity.

Fig. 2. Compares the reflection coefficient for both the Lorentz and Drude instances utilizing New ADE-TLM and accurate analytical results.

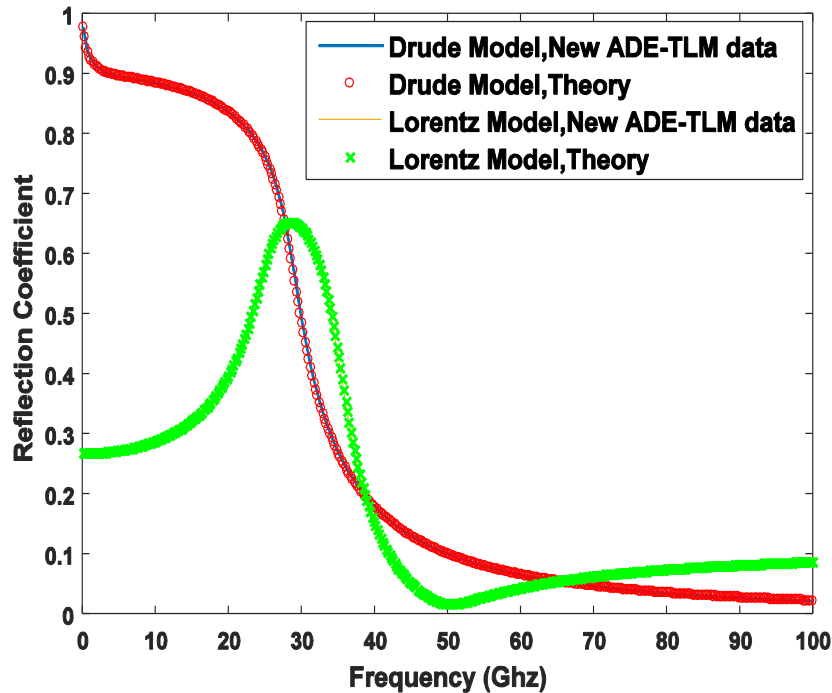


Fig. 2. Comparison of the new ADE-TLM and exact theoretical results from for the reflection coefficient for both the Lorentz and Drude half-space examples.

V. DISCUSSION

An efficient agreement was noted between the exact theoretical results as defined by Eq. (24) and the new ADE-TLM results for both the Lorentz and Drude examples. This result agrees very well [21].

VI. CONCLUSION

In this study, we provide a novel ADE-TLM approach for modeling Lorentz and Drude Dispersive media. This approach employs the connection between centered derivatives approximations, polarization current density J , and electrical voltage. In our technique, we used voltage sources that model linear characteristics and utilized the principle of variable admittance. The results obtained utilizing our New ADE-TLM method are in excellent agreement with the analytical values of the reflection coefficient demonstrating the validity of the proposed method.

REFERENCES

- [1] W.J.R. Hofer, «The Transmission-Line Matrix Method--Theory and Applications» IEEE Trans. MTT, vol 33, 882 – 893, Oct 1985
- [2] H. Jin and R. Vahldieck, « A frequency domain TLM method» IEEE IMSDA, pp 775-778, June 1-5.1992
- [3] Johns, P. B., “A symmetrical condensed node for the TLM method,” IEEE TMTT., vol 35, No. 4, pp 370–377, 1987
- [4] Hofer, J. R., “The transmission line matrix method, theory and applications,” IEEE TMTT., vol 33, No. 10, pp 882–893, 1985
- [5] M. Iben Yaich, M. Khalladi, I. Zekik, and J. Morente, “Modeling of frequency dependent magnetized plasma in hybrid Symmetrical condensed TLM method”, IEEE MWCL, vol 12, pp 293-295, 2002
- [6] S. E. Adraoui, M. Khalid, A. Zugari, M. Yaich, M. Khalladi, “Novel CDRC-TLM algorithm for the analysis of magnetized plasma”, Optik, vol 125, pp 276–279, 2014
- [7] S. E. Adraoui, E. S., A. Zugari, M. Bassouh, M. Yaich, and M. Khalladi, “Novel PLRC-TLM algorithm implementation for modeling electromagnetic wave propagation in gyromagnetic media”, IJAST., vol 6, No. 1, pp 26–32, 2013

- [8] K. Mounirh, S. El Adraoui, M. Charif, M. Iben Yaich, M. Khalladi, "Modeling of anisotropic magnetized plasma media using PLCDRC-TLM method", *Optik*, vol 126, pp 1479–1482, 2015
- [9] N. Doncov, T. Asenov, Z. Stankovic, J. Paul and B. Milovanovic, "TLM Z-Transform Method, Modelling of Lossy Grin MTM with Different Refractive Index Profiles", *SEE*, pp 103-112, August 2012
- [10] R. Abrini, M. Iben Yaich and M. Khalladi, "Efficient modeling of isotropic cold plasma media using JE-TLM method", *IEICE EE*, vol 4, No. 15, pp 492-497, 2007
- [11] K. Mounirh, S. El Adraoui, Y. Ekdiha, M.I. Yaich, M. Khalladi, "Modeling of dispersive chiral media using the ADE-TLM method," *PER.*, vol 64, pp 157–166, 2018
- [12] E. El ouardy, H. Bezzout, N. Meskini, and H. El Faylali, "Modeling of Drude Medium with Two New ADE-TLM Algorithms", *IJMOT.*, vol 18, NO. 4, pp 359–366, July. 2023
- [13] A. Attalhaoui¹, H. Bezzout, E. El Ouardy, M. Hanna, M. Habibi and H. El Faylali, "Modeling of Optical Pulse Propagation in Kerr and Raman Nonlinear Dispersive Media Using JE-TLM Method", *PIERM*, vol 72, pp 69-77, Aug. 2018
- [14] E. El ouardy and H. El Faylali, "The New ADE-TLM Algorithm for Modeling Debye Medium", *IJECES.*, vol 10, to be published
- [15] R. Holland, L. Simpson, and K. S. Kunz, "Finite-difference analysis of EMP coupling to lossy dielectric structures," *IEEE TEC*, vol 22, no. 3, pp 203–209, Aug. 1980
- [16] M. Okoniewski, M. Mrozowski, and M. A. Stuchly, "Simple treatment of multi-term dispersion in FDTD," *IEEE MGWL*, vol 7, no. 5, pp 121–123, May 1997
- [17] H. H. Abdullah, F. M. El Hefnawi, and A. Z. Elsherbeni, "A FDTD scattered field formulation for dispersive media," in *Proc. IEEE APSIS*, vol 1, pp 248–251, Jul. 2000
- [18] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Time Domain-Method*, 3rd ed. Norwood, MA: Artech House, 2005
- [19] A. Taflove and S.C. Hagness, "Computational Electrodynamics: The Finite-Difference Time-Domain Method", Artech House, Norwood, MA, 1995
- [20] S.J. Huang, "Exponential time differencing algorithm for Debye medium in FDTD", *Int. J. In. M. Wav.*, vol 28, pp 1025–1031, 2007
- [21] S.C. Kong, J. J. Simpson and V. Backman, "ADE-FDTD Scattered-Field Formulation for Dispersive Materials", *IEEE Microwave and Wireless Components Letters*, vol 18, pp 658–665, Jan. 2008
- [22] O. P. Gandhi, B. Q. Gao, J. Y. Chen, "A frequency-dependent finite difference time-domain formulation for general dispersive media", *IEEE Trans. Micr. Th. T.*, vol 41, pp 658–665, 1993

Створення тренувальних датасетів для корекційних великих мовних моделей

Нестеренко Володимир

*Кафедра радіофізики та комп'ютерних технологій
Львівський національний університет імені Івана Франка
Львів, Україна*

Анотація—У роботі розглянуто проблематику генерації наборів даних для дотренування великих мовних моделей для вирішення конкретного завдання. Перевірено раніше запропоновані підходи до вирішення цієї проблеми, що ґрунтуються на використанні open-source засобів та інструментів обробки природної мови.

Ключові слова—*Інженерія програмного забезпечення, Комп'ютерні науки, Машинне навчання, Великі мовні моделі, Обробка даних, Обчислювальна лінгвістика.*

I. ВСТУП

Серед великої кількості задач великих мовних моделей є виправлення помилок у тексті та спілкування у ролі помічника. Поєднання їх двох полягає у створенні чат-бота, який виправляє різного роду помилки у повідомленнях користувача. Такий програмний продукт є корисним при вивченні іноземної мови для зручного практичного застосування розмовної лексики та набуття необхідного досвіду. Розробка описаного програмного продукту потребує дотренування великої мовної моделі на відповідних структурованих даних, що представлені набором тренувальних пар [1]. Тренувальні пари у такому випадку складаються з повідомлення користувача та відповіді чат-бота. Генерація тренувального набору даних достатнього розміру та якості є одним з найважливіших етапів тренування моделей машинного навчання, тому в цій роботі буде перевірено кілька підходів до вирішення цього завдання та оцінено їх застосовність.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Актуальність великих мовних моделей в сучасному світі зростає з кожним днем. Це підтверджується популярністю моделі ChatGPT та дослідженнями від Google Cloud AI Research, OpenAI, Language Technologies Institute, тощо [2]. У деяких нещодавніх дослідженнях було також запропоновано різні методи зміни тексту з метою генерації, аугментації даних та захисту авторського права [3; 4; 5]. Наведені в них підходи засновані на використанні як простих алгоритмів так і статистичних підходів з машинним навчанням і покликані збільшити різноманітність текстів для подальшого використання. Аналіз на кількох мовних рівнях, а також інструменти створення векторних представлень, описані у розглянутих статтях, були використані для формування тренувальних наборів даних і вирішення завдання описаного в цій роботі.

III. МЕТОДИ

Мовний помічник проектується для німецької мови та з використанням великої мовної моделі «TheBloke/em_german_leo_mistral-GGUF». Враховуючи принципи роботи великих мовних моделей [6] та помилки, що найчастіше допускаються у текстах, для формування тренувального набору даних було використано кілька методів спотворення тексту: лематизація випадкового слова, зміна векторного представлення тексту, зміна порядку слів, заміна деяких слів та видалення випадкового слова. Наведені методи використовуються для

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

симуляції помилок, що може допускати людина при формулюванні повідомлення. Для отримання векторного представлення було використано інструмент `huggingface.AutoTokenizer` моделі «bert-base-german-cased». Спотворення векторного представлення тексту відбувалося шляхом зміни випадкового значення вектору на випадкову цілу величину, що описується розподілом Пуасона. Лематизація та зміна порядку слів з урахуванням синтаксичних залежностей були проведені за допомогою бібліотеки `spaCy` з моделлю «de_core_news_sm». Заміна слів проводилася для артиклів, прийменників та займенників з випадковим їх вибором. Міра схожості спотворених текстів на оригінальні обчислювалася шляхом отримання векторних представлень за допомогою інструменту `Sentence transformers` з моделлю «T-Systems-onsite/german-roberta-sentence-transformer-v2» та знаходженням оцінки схожості на основі їх косинусу подібності. У якості корпусу текстів було використано 52 діалоги німецькою мовою із навчального ресурсу з вільним доступом. Після їх обробки було отримано 1610 текстів, чого достатньо для попередньої перевірки гіпотези про застосовність кожного з методів спотворення текстів.

IV. РЕЗУЛЬТАТИ

Для описаних методів у результаті було отримано спотворені речення з такими середніми мірами схожості на оригінальні: лематизація – 0.994, спотворення векторних представлень – 0.842, перестановка – 0.99, заміна – 0.977, видалення – 0.945, де міра схожості 1 відповідає ідентичним реченням, а 0 – абсолютно несхожим.

V. ОБГОВОРЕННЯ

Отримані результати свідчать про те, що навіть незначна зміна векторного представлення тексту призводить до його суттєвої зміни. Водночас, при детальному аналізі спотворених речень, можна помітити, що зміни в реченнях є доволі далекими від тих помилок, що зазвичай роблять люди. Більш правдоподібні помилки отримуються за допомогою лематизації та заміни, хоча ці зміни і спричиняють набагато менші зміни в текстах. Варто зауважити, що результати дослідження не змушують нас відмовитися від будь-якого з методів, оскільки у тренувальних даних важливо навести як правдоподібні помилки, так і певного роду викиди та крайні випадки, аби модель після дотренування могла правильно реагувати на різні дані.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Аналіз результатів показує, що використання наведених методів спотворення даних потребує різної частоти застосування через різну міру спотворення, що вони вносять, та подібність до справжніх людських помилок. Частота та способи застосування методів спотворення може бути визначена статистичними методами, що і буде однією з тем подальших досліджень.

VII. ВИСНОВКИ

У ході проведення дослідження було з'ясовано, що представлені методи спотворення різною мірою впливають на текст, через що повинні використовуватися по-різному. Подальше формування тренувального набору даних для великої мовної моделі буде відбуватися з урахуванням міри спотворення кожного з методів зміни тексту.

ЛІТЕРАТУРА

- [1] Roshni Thoppilan et al., “LaMDA: Language Models for Dialog Applications,” Jan. 2022, doi: <https://doi.org/10.48550/arxiv.2201.08239>.
- [2] L. Floridi and M. Chiriatti, “GPT-3: Its Nature, Scope, Limits, and Consequences,” *Minds and Machines*, vol. 30, no. 4, pp. 681–694, Nov. 2020, doi: <https://doi.org/10.1007/s11023-020-09548-1>.
- [3] M. Bayer, M.-A. Kaufhold, and C. Reuter, “A Survey on Data Augmentation for Text Classification,” *ACM Computing Surveys*, p. 3544558, Jun. 2022, doi: <https://doi.org/10.1145/3544558>.
- [4] Efsthios Stamatatos, “Authorship Attribution Using Text Distortion,” Jan. 2017, doi: <https://doi.org/10.18653/v1/e17-1107>.

- [5] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 8, no. 1, Jul. 2021, doi: <https://doi.org/10.1186/s40537-021-00492-0>.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

Використання машинного навчання для управління процесами аграрної економіки

Петро Грицюк, Максим Гаврилук

*Кафедра комп'ютерних технологій та економічної кібернетики
Національний університет водного господарства та природокористування,
Рівне, Україна*

Анотація—Врожайність пшениці в Україні в останні двадцять років значно зросла завдяки значним інвестиційним надходженням та покращенню технології вирощування, збирання та зберігання врожаю. Разом з тим, врожайність зазнає значних коливань, які зумовлені впливом кліматичних факторів. В роботі проведено дослідження впливу кліматичних факторів на коливання врожайності озимої пшениці у різних регіонах України. Інструментом досліджень виступали методи машинного навчання. Для реалізації даної методики була проведена бінарзація значень врожайності.

Ключові слова—Врожайність пшениці, Кліматичні фактори, Машинне навчання, Бінарзація врожайності, Класифікація.

I. ВСТУП

Використання машинного навчання в аграрній економіці відкриває широкі можливості для оптимізації та управління різноманітними аспектами сільськогосподарської діяльності. Моделі машинного навчання можуть аналізувати великі обсяги даних про ґрунти, кліматичні умови та інші фактори, що впливають на урожайність. Це дозволяє сільгоспвиробникам зробити більш точні прогнози та вжити необхідні заходи для зменшення ризиків.

Виробництво зерна є однією з найважливіших галузей економіки України, що забезпечує продовольчі потреби населення та стабільний приплив валюти. Середньорічне виробництво зернових в Україні за 2019-2021 роки досягло рівня 75 млн тон, а середньорічний експорт за цей час склав 50 млн тон. Основною зерною культурою в Україні є пшениця. Середньорічне виробництво пшениці в Україні за 2019-2021 роки досягло рівня 26,5 млн тон [1].

З точки зору природно-кліматичних характеристик, які впливають на вирощування зернових культур, територію України умовно можна розділити на три великих зони : степова зона (Кіровоградська, Дніпропетровська, Запорізька, Херсонська, Миколаївська, Одеська області), лісостепова чорноземна зона (Чернігівська, Сумська, Полтавська, Харківська, Київська, Черкаська, Вінницька області) та західний регіон (Волинська, Рівненська, Житомирська, Львівська, Івано-Франківська, Тернопільська та Хмельницька області). Закарпатська та Чернівецька області мають відмінні кліматичні характеристики і не відносяться до жодного з названих вище кластерів.

Головною передумовою хорошої врожайності пшениці є її успішна вегетація у весняно-літній період. На кожному етапі вегетації існує вікно оптимальних значень кліматичних факторів. Вихід факторів за межі цього вікна приводить до зниження врожайності. Для аналізу та моделювання впливу кліматичних факторів на коливання врожайності пшениці ми використали масив даних за період 2000 – 2021 роки. Ми використовували наступні кліматичні фактори : середньодекадна температура за квітень, травень та червень та місячні суми опадів за ці ж місяці.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Виробництво пшениці є основою українського сільського господарства, але зміни клімату загрожують йому в деяких регіонах України. Багато досліджень останніх років присвячено дослідженню впливу кліматичних факторів на врожайність сільськогосподарських культур. Вплив зміни клімату на врожайність озимої пшениці в Україні та окремо для трьох окремих екологічних зон України було оцінено у великому аналітичному огляді, проведеному в рамках проекту «Німецько-український аграрно-політичний діалог» [2]. За висновком авторів, головне занепокоєння викликає родюча степова зона, де клімат більш жаркий і сухий, а також спостерігаються часті посухи. Майбутні зміни клімату можуть посилити ці негативні наслідки та спричинити опустелювання цього регіону.

У роботі [3] аналізується вплив клімату на річну мінливість урожайності основних сільськогосподарських культур у США. Результати аналізу свідчать про те, що кліматичні зміни впливають на рівень і коливання врожайності сільськогосподарських культур. Збільшення частоти посух в останні роки розглядається як головна загроза для сільського господарства на роботі [4]. Автор дослідив вплив зміни клімату на рівень виробництва основних сільськогосподарських культур, а також на ВВП Угорщини.

Науково-технічний прогрес сприяв надходженню великих обсягів статистичних даних з різних галузей сільського господарства. Це значно розширило можливості використання комп'ютерних технологій для аналізу та моделювання кліматичних впливів на врожайність сільськогосподарських культур. Поява технології великих даних призвела до появи нових потужних аналітичних інструментів, таких як методи машинного навчання, які раніше позитивно зарекомендували себе в медицині, фінансах і біології [5]. В останні роки з'явилися публікації, в яких описується застосування методів машинного навчання для прогнозування врожайності сільськогосподарських культур. Аналітичні моделі, такі як модель дерева рішень, модель випадкового лісу, машина опорних векторів, байєсовські мережі та штучні нейронні мережі, використовуються для аналізу ключових факторів, що впливають на врожайність. Ці методи дають змогу аналізувати ґрунтовий, кліматичний і водний режими, які суттєво впливають на ріст і врожайність культур. В огляді [6] представлені підходи машинного навчання до вирішення цієї задачі з точки зору прикладного економіста.

У дослідженні [7] запропоновано систему прогнозування на основі машинного навчання для передбачення врожайності шести сільськогосподарських культур у країнах Західної Африки. Для побудови такої системи були використані дерева рішень, багатовимірні логістична регресія та модель найближчих сусідів. Було виявлено, що результати прогнозування моделі дерева рішень і моделі k -Nearest Neighbor корелюють з очікуваними даними.

Структура глибокого навчання для прогнозування врожайності з використанням даних дистанційного зондування представлена в роботі [8]. Запропоновано підхід до зменшення розмірності на основі гістограм та продемонстровано структуру глибокого гаусового процесу, за допомогою якого усуваються просторово корельовані похибки та значно підвищується точність прогнозування врожайності сої в графствах США.

Одним із найпотужніших інструментів машинного навчання є штучні нейронні мережі. У статті [9] використовується напівпараметричний варіант глибокої нейронної мережі, яка може одночасно враховувати складні нелінійні зв'язки у масивах даних великої розмірності. Використовуючи дані про врожайність кукурудзи на Середньому Заході США, було показано, що при прогнозуванні врожайності цей підхід перевершує як класичні статистичні методи, так і непараметричні нейронні мережі. У роботах [10-12] також використовувалися моделі штучної нейронної мережі для моделювання динаміки врожайності.

III. МЕТОДИ

Аналіз динаміки врожайності пшениці в регіонах України за останні 22 роки показує, що врожайність зростає [13]. Підвищення врожайності стало наслідком підвищення інвестиційної

привабливості зернової галузі та значних інвестицій, які надійшли в галузь. В результаті покращилася насіннева база, підвищилась агротехнічна культура та розвинулася логістична мережа (елеватори, зерновози, порти). Для моделювання тренду врожайності tr_t ми використали лінійну модель

$$tr_t = a_0 + ta_1. \quad (1)$$

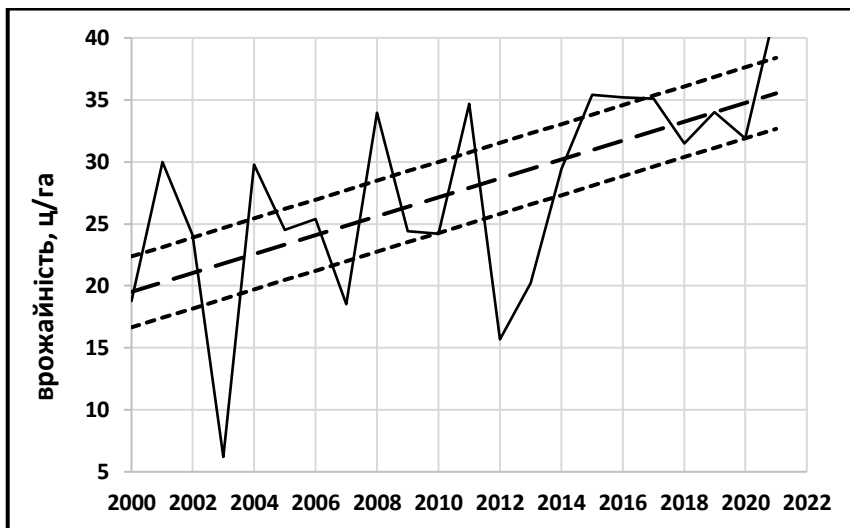


Рис. 1. Динаміка врожайності пшениці у Херсонській області. Штрихова лінія зображає лінійний тренд. Пунктирні лінії зображують межі низької та високої врожайності. Розраховано авторами за даними [1]

Проте тенденція до підвищення врожайності y_t супроводжується значними коливаннями врожайності, причиною яких здебільшого є вплив погодно-кліматичних факторів. Ці коливання є відхиленнями від тренду eps

$$eps_t = y_t - tr_t. \quad (2)$$

Для моделювання компоненти врожайності eps , яка відображає вплив клімату, ми використали статистику врожайності пшениці та кліматичні дані для областей України за період 2000 – 2021 роки. Кліматичні характеристики взяті нами з ресурсу [14], дані про врожайність взяті з ресурсу [1]. Вирішальний вплив на врожайність пшениці має її успішна вегетація в період з квітня по червень. Для оцінки впливу клімату на врожайність пшениці ми використовували середньодекадні значення температури квітня, травня, червня та місячні суми опадів за цей період. Урожай пшениці формується під впливом 12 кліматичних факторів (9 температурних і 3 пов'язаних з опадами). При цьому відгук eps пов'язаний через рівняння множинної регресії з факторними ознаками $t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, R_{10}, R_{20}, R_{30}$. Модель залежності відхилення врожайності eps від кліматичних факторів має вигляд:

$$eps = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_3 t_3 + \beta_4 t_4 + \beta_5 t_5 + \beta_6 t_6 + \beta_7 t_7 + \beta_8 t_8 + \beta_9 t_9 + \beta_{10} R_{10} + \beta_{20} R_{20} + \beta_{30} R_{30} + \varepsilon. \quad (3)$$

Однією з перешкод для використання машинного навчання в задачах сезонного прогнозування є обмежений розмір вибірки даних спостережень. Щоб уникнути цієї проблеми, ми реалізували модельне навчання на об'єднаних ансамблях вибірок, які поєднують дані з сусідніх областей із подібними погодними та кліматичними умовами та типами ґрунтів [15]. Ми отримали такі вибірки, об'єднуючи області України у агрокліматичні зони, описані вище. Відповідні вибірки даних будуть містити від 132 до 154 зразків, кожен з яких містить детрендовану врожайність eps і 12 кліматичних факторів.

IV. РЕЗУЛЬТАТИ

Для обробки масивів даних ми використали середовище програмування R із розширенням RStudio. Кореляційно-регресійний аналіз дозволив виявити ті фактори, які відіграють вирішальний вплив на зміни врожайності пшениці (табл. 1). Знак «+» позначає позитивний вплив фактора, знак «-» - негативний вплив фактора.

Таблиця 1. Вплив кліматичних факторів на врожайність пшениці

	<i>t1</i>	<i>t2</i>	<i>t3</i>	<i>t4</i>	<i>t5</i>	<i>t6</i>	<i>t7</i>	<i>t8</i>	<i>t9</i>	<i>R10</i>	<i>R20</i>	<i>R30</i>
Степ					-		-			+		
Лісостеп			+	-			-					
Захід					-							-

Для вирішення багатьох проблем при плануванні аграрного бізнесу, немає необхідності мати точний прогноз врожайності. Для успішного управління фінансовим ризиком важливо передбачити негативне відхилення від тренду, тобто низьке значення врожайності. Тому головним завданням даного дослідження є прогнозування низьких показників врожайності пшениці. До групи «низька врожайність» ми відносимо ті значення врожайності, для яких виконується умова $F(eps) < 0.33$. Тут $F(x)$ – інтегральна функція нормального розподілу трендових залишків врожайності eps .

Для реалізації класифікаційного підходу до передбачення врожайності, ми вводимо двійкову змінну $eps1$, яка має лише два значення: 1 («низька врожайність») і 0 («висока врожайність») [13]. Значення змінної $eps1$ визначається за правилом

$$eps1 = \begin{cases} 1, & \text{if } F(eps) < 0.33; \\ 0, & \text{if } F(eps) \geq 0.33. \end{cases} \quad (3)$$

Для категоріального прогнозування врожайності пшениці ми використали декілька алгоритмів класифікації: LDA – лінійний дискримінантний аналіз, SVL – метод опорних векторів (лінійний базис), SVR – метод опорних векторів (радіальний базис), CART – метод дерева рішень, RF – метод випадкового лісу, GLM – метод логістичної регресії. Для оцінки продуктивності побудованих нами класифікаторів використовуються наступні показники: матриця неточностей (*confusion matrix*), точність, чутливість, специфічність і AUC - площа під кривою ROC (*receiver operating characteristic*). Площа під кривою ROC є універсальним критерієм оцінки класифікатора [16].

Для порівняння ефективності різних методів машинного навчання ми використали методику, описану у [17]. У якості ілюстрації нижче наведені результати прогностичної класифікації врожайності пшениці з використанням методів машинного навчання для степового регіону (табл. 2). Серед використаних нами методів найвищу точність (84%) продемонстрували метод випадкового лісу та метод опорних векторів.

Таблиця 2

Критерії якості методів машинного навчання

	Accuracy	Sensitivity	Sensitivity	AUC
LDA	0.812	0.846	0.667	0.853
SVL	0.844	0.828	1.000	0.853
SVR	0.844	0.828	1.000	0.993
CART	0.781	0.840	0.571	0.749
RF	0.844	0.828	1.000	0.987
GLM	0.812	0.846	0.667	0.987

V. ВИСНОВКИ

Прогнозування врожайності із врахуванням впливу кліматичних факторів є ефективним інструментом управління збиранням, зберіганням та наступною логістикою врожаю сільськогосподарських культур в умовах зміни клімату. Побудовані нами моделі дозволяють заздалегідь (за 3 місяці) оцінити майбутню врожайність, що дозволяє оптимізувати інвестиційні та маркетингові рішення. Запропоновану нами методику можна розширити для класифікації врожайності на три категорії: «висока врожайність», «середня врожайність» і «низька врожайність». Цей інструмент дозволить керувати ризиками та збільшувати прибутковість зерновиробництва. Таким чином, прогнозування врожайності за допомогою машинного навчання є важливим інструментом ефективного управління процесами аграрної економіки.

В інших наших роботах ми досліджували вплив кліматичних факторів на коливання врожайності пшениці в чорноземній зоні України та у західному регіоні України. Результати цих досліджень зараз перебувають у друці. Таким чином, ми застосували методику машинного навчання до вивчення впливу клімату на коливання врожайності пшениці в різних регіонах України. Наші дослідження є внеском у вирішення проблеми забезпечення сталості виробництва зерна в Україні. Запропонована нами методика може бути поширена на дослідження впливу клімату на сільськогосподарські культури в різних регіонах і країнах.

ЛІТЕРАТУРА

- [1] Державна служба статистики України [Електронний ресурс], Режим доступу: <https://www.ukrstat.gov.ua/>
- [2] D. Muller, A. Jungandreas, F. Koch, F. Shirhorn, “The impact of climate change on wheat production in Ukraine”. Report on agricultural policy (APD), 2016
- [3] C.C. Chen, B.A. McCarl, D.E. Schimmelpfennig “Yield Variability as Influenced by Climate: A Statistical Investigation”, *Clim. Change*, vol. 66, pp. 239–261, 2004 DOI <https://doi.org/10.1023/B:CLIM.0000043159.33816.e5>
- [4] H.M. Zemankovics, “Mitigation and adaptation to Climate Change in Hungary”, *J. Central Eur. Agric.*, Vol. 13, no.1, pp. 58-72, 2012 DOI: <https://doi.org/10.5513/JCEA01/13.1.1015>
- [5] Jake VanderPlas, *Python Data Science Handbook. Essential Tools for Working with Data*: Sebastopol, O’Reilly Media, 2016
- [6] H. Storm, K. Baylis, T. Heckelei, “Machine learning in agricultural and applied economics”, *Eur. Rev. Agric. Econ.*, Vol. 47, no. 3, pp. 849–892, 2020 <https://doi.org/10.1093/erae/jbz033>
- [7] L.S. Cedric, W.Y. Hamilton Adoni, R. Aworkaa, J.T. Zoueuud, F.K. Mutombo, M. Krichen et al., “Crops yield prediction based on machine learning models: Case of West African countries”, *Sm. Agri. Tech.*, Vol. 2, pp. 1–14, 2022 DOI: [10.1016/j.atech.2022.100049](https://doi.org/10.1016/j.atech.2022.100049)
- [8] J. You, X. Li, M. Low, D. Lobell, S. Ermon, “Deep gaussian process for crop yield prediction based on remote sensing data”, *The Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, Vol. 31, no. 1, pp. 4559-4565 DOI: <https://doi.org/10.1609/aaai.v31i1.11172>
- [9] A. Crane-Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture”, *Environ. Res. Lett.*, Vol. 13, no. 11, pp. 1–12, 2018 <https://doi.org/10.1088/1748-9326/aae159>
- [10] T. Klompenburg van, A. Kassahun, C. Catal, “Crop yield prediction using machine learning: a systematic literature review”, *Comput. Electron. Agric.*, Vol. 177, 2020 <https://doi.org/10.1016/j.compag.2020.105709>
- [11] S. Khaki, L. Wang, “Crop yield prediction using deep neural networks”, *Front. Plant Sci.*, Vol. 10., 22 May 2019, 2019 <https://doi.org/10.3389/fpls.2019.00621>
- [12] R. Tanabe, T. Matsui, T.S. Tanaka, “Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery”, *Field Crops Res.*, Vol. 291, 2023 <https://doi.org/10.1016/j.fcr.2022.108786>
- [13] П.М. Грицюк, Т.Ю. Бабич, Б.В. Красько, “Класифікаційні методи прогнозування врожайності”, *Вісник Хмельницького національного університету. Технічні науки*, том 309, №3, с. 209 – 216, 2022 <https://www.doi.org/10.31891/2307-5732-2022-309-3-209-210>
- [14] Архів метеоданих. [Електронний ресурс] – Режим доступу: <https://meteopost.com/weather/archive/>
- [15] P. Gibson, W. Chapman, A. Altinok et al., “Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts”, *Communications Earth & Environment*, vol. 2, no. 159, 2021 <https://doi.org/10.1038/s43247-021-00225-4>
- [16] T. Fawcett, “An introduction to ROC analysis”, *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006 <https://doi.org/10.1016/j.patrec.2005.10.010>
- [17] M. Kuhn, K. Johnson, *Applied Predictive Modeling*. New York: Springer, New York, 2013

Identify the perpetrator in the gathering Using AI

Soumya Upadhyay¹

Department of Cyber Security
COER University, Roorkee
ersoumya26@gmail.com

Vishal Chauhan²

Dept. of Electrical & Electronics
COER University, Roorkee
vishalchauhan.itg@gmail.com

Vaibhav Sharma³

Department of Cyber Security
COER University, Roorkee
vaibhavsharma9979@gmail.com

Yashveer Makhloga⁴

Department of Cyber Security
COER University, Roorkee
yashveermakhloga0@gmail.com

Madhav⁵

Dept. of Electrical & Electronics
COER University, Roorkee
madhavdua26@gmail.com

Abstract—This paper provides a novel method for improving public safety and security by leveraging smart eyewear with built-in cameras for real-time analysis of people in crowded places. The system uses advanced face recognition algorithms and artificial intelligence to detect people with prior criminal records in a crowd, allowing for a proactive reaction to possible security risks. The technology ensures continuous surveillance without infringing on private rights by including strong data protection mechanisms as well as ethical concerns. The system's accuracy and dependability are carefully verified by combining cutting-edge deep learning techniques and varied datasets. The findings of this study have important implications for law enforcement, enabling proactive measures to prevent criminal activity while also contributing to the ongoing debate about balancing security needs with individual privacy rights in the responsible use of emerging surveillance technologies.

Keywords—ML-Machine Learning, AI-Artificial Intelligence, DL-Deep Learning.

I.INTRODUCTION

The confluence of artificial intelligence and surveillance technologies has altered public safety paradigms in an era of rapid technology breakthroughs. One such game-changing breakthrough is the incorporation of smart glasses with miniature built-in cameras, which presents a disruptive approach for increasing security in congested areas. The primary goal of this research is to use the capabilities of these smart glasses to undertake real-time analysis of persons within dense crowds, with a particular emphasis on identifying those with prior criminal records. Traditional methods of criminal identification and crowd surveillance have frequently proved inefficient, time-consuming, and ineffective. These smart glasses give law enforcement and security officers unparalleled capabilities by utilizing cutting-edge facial recognition algorithms and deep learning techniques[1]. The device can assess large crowds in a covert and inconspicuous manner, instantaneously comparing facial features against a comprehensive criminal database. This rapid identification of people with a history of criminal activity allows for pre-emptive actions, reducing potential risks before they escalate. Furthermore, the incorporation of ethical considerations and strong data protection standards ensures that individual privacy rights are honoured, addressing common worries about such advanced monitoring technologies[2]. This project not only marks a substantial advancement in public safety, but it also adds to the continuing discussion about the appropriate and accountable use of developing technologies. This research, by diving into the complicated

balance between security imperatives and civil rights, places it at the forefront of new solutions, paving the way for a future in which smart glasses play a critical role in guaranteeing the safety and security of our communities.

II.LITERATURE REVIEW

Advances in surveillance technologies, particularly those based on artificial intelligence (AI) and facial recognition algorithms, have spurred heated debates about public safety, privacy, and ethical concerns in recent years. The introduction of wearable gadgets such as smart glasses with miniature built-in cameras has provided new opportunities for law enforcement authorities to improve security measures, particularly in congested areas. This literature review investigates the present studies, obstacles, and ethical concerns about the use of smart glasses in densely populated regions to analyse individuals with prior criminal records[2].

Smart Glasses Technological Advancements

Smart glasses with miniature cameras have made significant advances in their capabilities, making them into sophisticated tools for real-time data analysis. These gadgets have high-resolution cameras, microprocessors, and connection capabilities, allowing for simple integration with facial recognition software[3]. Researchers have proved the use of smart glasses in a variety of applications, ranging from assisting visually impaired people to improving situational awareness for military troops.

Technology for Facial Recognition

Deep learning algorithms have aided in the advancement of facial recognition technology, a subset of AI. These techniques, notably convolutional neural networks (CNNs), have demonstrated outstanding accuracy in detecting and recognizing faces in photos and videos[4]. Many research have looked into the use of facial recognition in law enforcement, from identifying suspects in criminal investigations to monitoring public spaces for security.

Surveillance and Public Safety

Public safety in congested areas is a serious concern for law enforcement authorities all around the world. Traditional surveillance technologies, such as closed-circuit television (CCTV) cameras, have real-time analysis and accuracy constraints. Smart glasses with face recognition capabilities take a more proactive approach, allowing for the quick identification of people with criminal backgrounds. This is especially important in areas like airports, train stations, and large-scale events where a big number of individuals must be analysed quickly.

Ethical Issues and Difficulties

While there are obvious benefits to using smart glasses for criminal record analysis, various obstacles and ethical considerations must be addressed. One important source of concern is the accuracy of facial recognition algorithms, which have been shown to contain biases, particularly regarding gender and ethnicity[3]. Researchers and politicians are working hard to reduce these biases in order to ensure fair and unbiased identification processes. Concerns about privacy also loom prominently in the context of facial recognition technologies. Individuals are always being watched, which raises concerns about their right to privacy and the potential misuse of personal data[5]. Balancing public safety and private privacy rights is a critical issue that governments and technology must address. To handle these challenges appropriately, ethical frameworks, legal rules, and transparent methods are required[4].

III.METHODOLOGY

The methodology in this research comprises using the Haar Cascade algorithm to match individuals' faces taken by smart glasses with photographs saved in a cloud server database. To ensure precision and dependability in recognizing persons with criminal histories among a crowd, the process of matching faces goes through multiple sophisticated processes.

The Haar Cascade technique is used for face detection which is an important phase in the process[4]. Haar features, which are essentially filters, are used to recognize facial features such as the eyes, nose, and mouth. To distinguish between faces and non-faces in images, these attributes are fed into a cascade classifier. The system examines the input image at multiple sizes with a sliding window technique, detecting probable face regions based on the trained attributes.

Once a face is detected, numerous facial traits like as the distance between the eyes, the curvature of the lips, and the shape of the nose are retrieved. These characteristics are then encoded into a distinct numerical representation, frequently employing techniques such as Local Binary Patterns (LBP) or Histogram of Oriented Gradients (HOG). This encoding converts the facial traits into a format that can be compared and analysed. To quantify the similarity of facial feature encodings, matching techniques such as Euclidean distance or Cosine similarity are commonly used.

To assess the level of similarity required for a match, a preset threshold is set. A notification is generated if the calculated similarity score exceeds this level, indicating a substantial match between the detected face and a record in the database. This notification is sent as a text message to the associated mobile device, alerting authorities to the identified individual with a criminal record. Additional tests and verifications can be added to reduce false positives. These could include taking numerous photographs of the recognized face from various angles, or using additional biometric indicators such as iris recognition to improve identification accuracy.

Concurrently, the smart glass gadget connects to the cloud server, retrieving the pre-processed face images and related encodings from the criminal records database. The encoded facial features of the smart glass-detected face are then compared to the encodings of individuals in the database. Fine-tuning algorithm parameters and continual training on varied datasets are also important in lowering false positives. This methodology ensures precise matching of individuals in busy areas by applying the Haar Cascade algorithm for face identification and establishing a robust comparison mechanism, permitting quick and accurate warnings to law enforcement agencies for preventive security steps. The project's heart is a complete approach to facial recognition that enables rapid and reliable detection of persons with criminal records utilizing smart glasses linked to a cloud server.

Libraries Used:

Face Recognition
PushBullet
OpenCv

Steps those have been followed:

1. Capture real time video through the camera
2. Encode the images of dataset before storing in a 2D array
3. Analyse the faces in real time video through Open Cv.
4. Match the features with the help of face recognition library from the available dataset.
5. On identifying the match; trigger a beep and display the name of the suspect found
6. Using the pushbullet library and API call, it will generate a notification to the mobile device to ping the authorities regarding the match with the details of the suspect found.



Fig:1- Assembles Eye Wear Glass with Wi-Fi Enables Camera

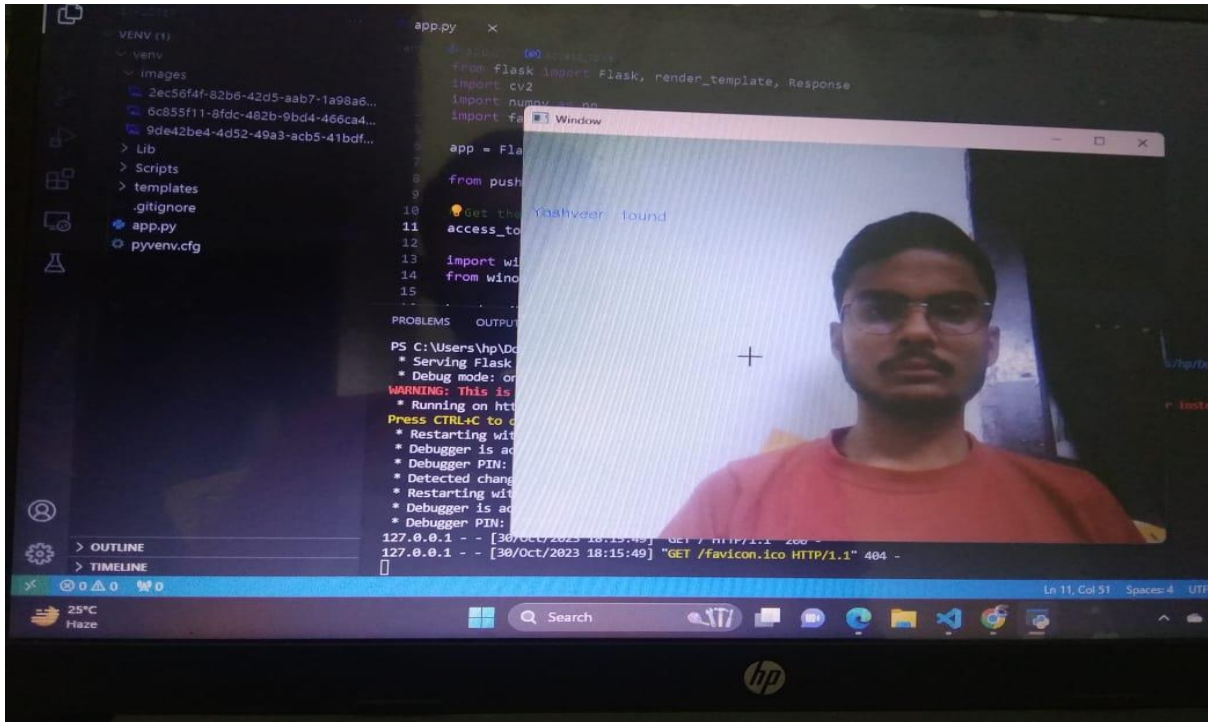


Fig:2- Capture real time video through the camera

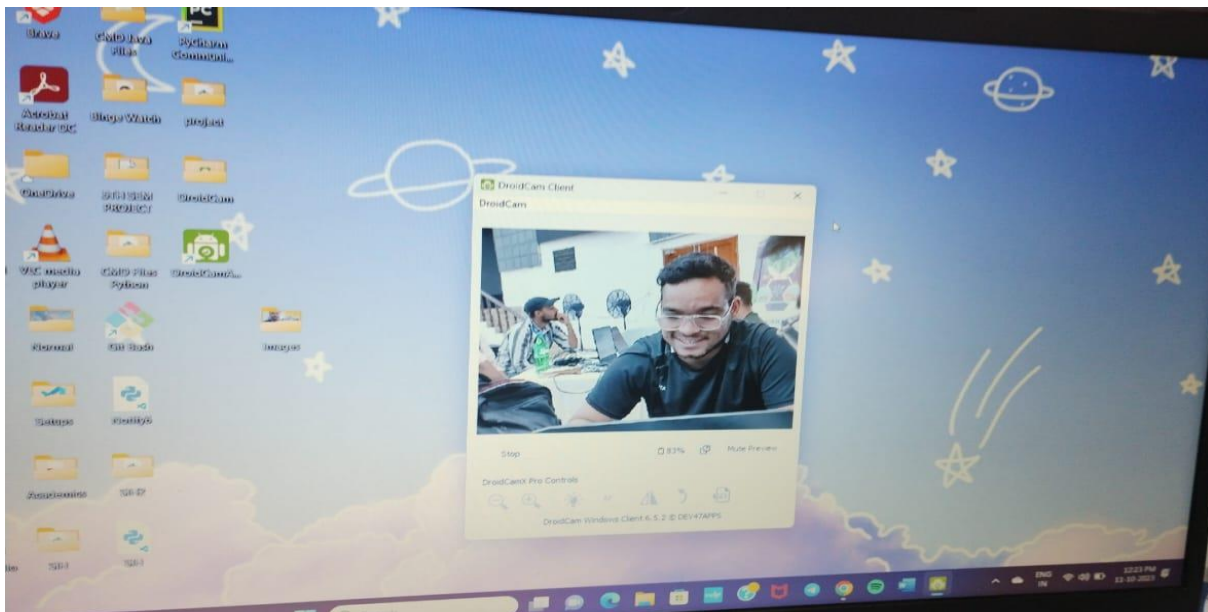


Fig:3- Tested model using the web cam

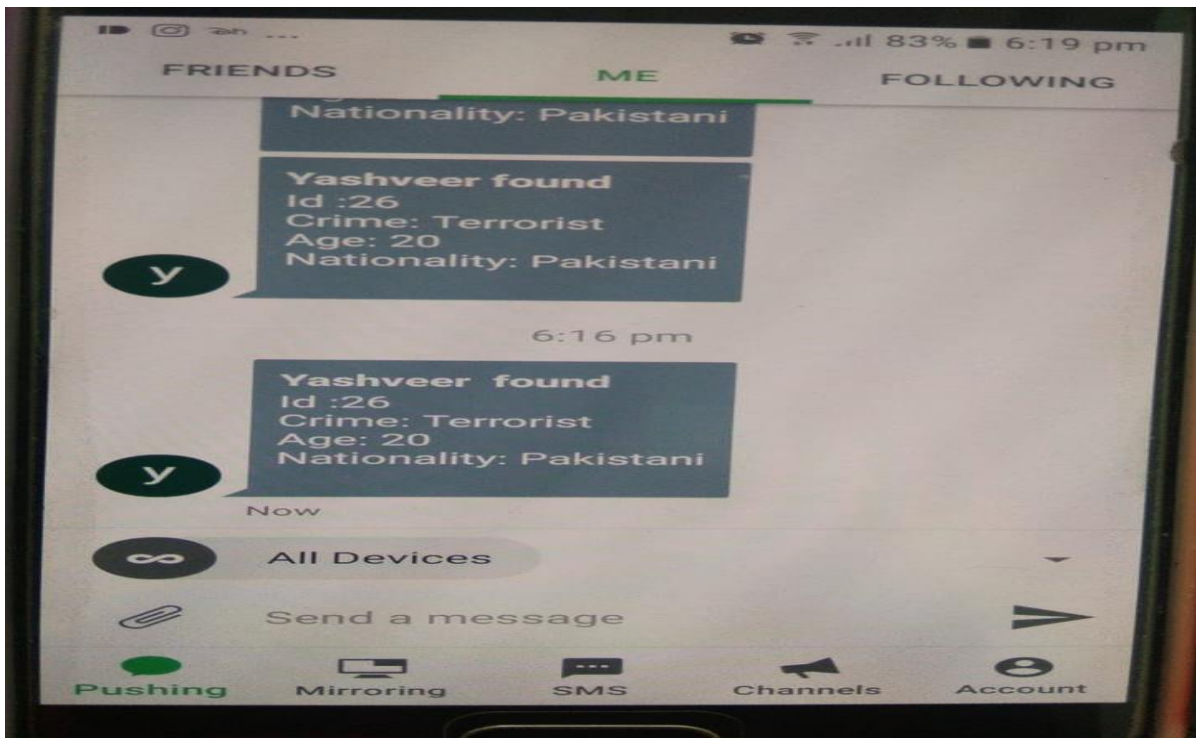


Fig:4- Generated Notification to the mobile

IV.RESULTS

Since no prior research of a similar nature has been located, the study is extremely recent. This research will assist law enforcement officials in identifying offenders and suspects in congested areas. The nation's safety will be enhanced by it. Our accuracy is 98.89% using Open CV and Pushbullet libraries.

V.CONCLUSION

Finally, with the creative integration of smart glasses, facial recognition technology and cloud computing, this paper marks a significant step toward improving public safety and security in congested areas. The system offers a strong way for identifying persons with criminal records by applying the Haar Cascade algorithm for exact face identification and comparison. The continuous connection to a cloud server, which allows for real-time database retrieval and matching, assures fast and accurate results. The study not only demonstrates the promise of developing technology in law enforcement, but it also stresses the significance of responsible deployment by addressing issues like as accuracy, bias, and privacy concerns. This initiative serves as a pioneering example of utilizing new technologies for the greater good of society, providing a pattern for future advancements in the field of public safety and surveillance by striking a balance between security imperatives and individual rights.

VI.DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

REFERENCES

- [1]. Kim, Dawon, and Yosoon Choi. 2021. "Applications of Smart Glasses in Applied Sciences: A Systematic Review" *Applied Sciences* 11, no. 11: 4956. <https://doi.org/10.3390/app11114956>.
- [2]. Davis, F.D.. (1989). Perceived usefulness, perceived ease, and user acceptance of information technology. *MIS Quarterly*. 13..
- [3]. L. -H. Lee and P. Hui, "Interaction Methods for Smart Glasses: A Survey," in *IEEE Access*, vol. 6, pp. 28712-28732, 2018, doi: 10.1109/ACCESS.2018.2831081.
- [4]. Zuidhof, N., Ben Allouch, S., Peters, O. et al. Defining Smart Glasses: A Rapid Review of State-of-the-Art Perspectives and Future Challenges From a Social Sciences' Perspective. *Augment Hum Res* 6, 15 (2021). <https://doi.org/10.1007/s41133-021-00053-3>
- [5]. M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 116-119, doi: 10.1109/ICCCIS48478.2019.8974493.
- [6]. A. Jain, A. Bholra, S. Upadhyay, A. Singh, D. Kumar and A. Jain, "Secure and Smart Trolley Shopping System based on IoT Module," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 2243-2247, doi: 10.1109/IC3I56241.2022.10073159.

Застосування Grasshopper Optimization Algorithm при навчанні нейронних мереж

Андрій Ляшкевич¹, Анастасія Дейнеко², Юлія Шевчук¹

¹ Кафедра системного аналізу та теорії прийняття рішень
Київський національний університет імені Тараса Шевченка

Київ, Україна

² ІТ СТЕП Університет

Львів, Україна

Анотація— У даній роботі пропонується метод розв'язання проблеми пошуку глобальних та локальних екстремумів функції помилки при навчанні нейронних мереж, який базується на використанні модифікації *Grasshopper Optimization Algorithm* для *Forward-Forward Neural Network*. Також наведено покровий алгоритм виконання даного методу.

Ключові слова — *Grasshopper Optimization Algorithm, Neural Network, Neural Network Optimizer, Forward-Forward Neural Network*.

I. Вступ

Процес знаходження найкращих значень змінних певної задачі для мінімізації або максимізації цільової функції називається оптимізацією. Оптимізаційні задачі існують у різних галузях знань. Щоб розв'язати оптимізаційну задачу, потрібно зробити кілька кроків. По-перше, слід визначити параметри задачі. Залежно від природи параметрів задачі можна класифікувати на неперервні та дискретні. По-друге, необхідно розпізнати обмеження, які накладаються на параметри [1]. Обмеження поділяють оптимізаційні задачі на обмежені та необмежені. По-третє, необхідно дослідити та врахувати цілі даної задачі. У цьому випадку оптимізаційні задачі класифікуються на одноцільові та багатоцільові [2]. Нарешті, на основі визначених типів параметрів, обмежень та кількості цілей слід обрати відповідний оптимізатор, який буде використовуватися для розв'язання задачі.

Математична оптимізація в основному спирається на градієнтне формування відповідних функцій для знаходження оптимального рішення. Хоча такі методи все ще використовуються різними дослідниками, вони мають певні недоліки. Підходи математичної оптимізації страждають від захоплення локальних оптимумів. Це означає, що алгоритм припускає, що локальний розв'язок є глобальним розв'язком, і таким чином не може отримати глобальний оптимум. Вони також часто є неефективними для задач з невідомими або обчислювально дорогими розв'язками похідних [3]. Іншим типом алгоритму оптимізації, який усуває ці два недоліки, є стохастична оптимізація [4].

Стохастичні методи покладаються на випадкові оператори, які дозволяють їм уникати локальних оптимумів. Всі вони починають процес оптимізації зі створення одного або набору випадкових розв'язків задачі. На відміну від математичних методів оптимізації, їм не потрібно обчислювати градієнт розв'язку, вони просто оцінюють розв'язки за допомогою цільової функції (функцій). Рішення про те, як покращити розв'язки, приймаються на основі розрахованих цільових значень. Таким чином, проблема розглядається як чорний ящик, що є дуже корисним механізмом при розв'язанні проблем з невідомими просторами пошуку. Завдяки цим перевагам методи стохастичної оптимізації стали дуже популярними протягом останніх двох десятиліть [5].

Серед підходів до стохастичної оптимізації найбільш популярними є популяційні алгоритми, натхненні природою [6]. Такі технології імітують природні методи розв'язання проблем, часто ті, що використовуються живими істотами. Виживання - головна мета для всіх істот. Щоб досягти цієї мети, вони еволюціонують і пристосовуються різними способами. Тому розумно шукати натхнення у природи як найкращого і найстарішого оптимізатора на планеті. Такі алгоритми поділяються на дві основні групи: на основі одного рішення та на основі декількох рішень. У першому класі для конкретної задачі генерується і вдосконалюється єдиний випадковий розв'язок. У другому класі, навпаки, для задачі генерується і покращується декілька розв'язків. Алгоритми на основі декількох рішень є більш популярними, ніж алгоритми на основі одного рішення, як показує література [7].

Алгоритми, що базуються на декількох розв'язках, за своєю суттю мають вищий рівень уникнення локальних оптимумів завдяки покращенню множинних рішень під час оптимізації. У цьому випадку розв'язок, що потрапив у пастку в локальному оптимумі, може отримати допомогу від інших розв'язків, щоб вискочити з локального оптимуму.

Багаторазові рішення досліджують більшу частину простору пошуку порівняно з алгоритмами, що базуються на одному рішенні, тому ймовірність знаходження глобального оптимуму є високою. Крім того, інформація про простір пошуку може обмінюватися між декількома рішеннями, що призводить до швидкого руху до оптимуму. Хоча алгоритми на основі декількох розв'язків мають ряд переваг, вони вимагають більшої кількості оцінок функцій.

II. Grasshopper Optimization Algorithm (GOA)

Найпопулярнішими алгоритмами на основі одного рішення є hill climbing [8] та simulated annealing [9]. Обидва алгоритми дотримуються схожої ідеї, але уникнення локальних оптимумів у SA є вищим через стохастичний фактор охолодження. Найпопулярнішими алгоритмами на основі декількох рішень є Genetic Algorithms (GA) [10], Particle Swarm Optimization (PSO) [11], Ant Colony Optimisation (ACO) [12], and Differential Evolution (DE) [13]. Алгоритм GA був натхненний дарвінівською теорією еволюції. У цьому алгоритмі розв'язки розглядаються як особини, а параметри розв'язків замінюють їхні гени. Виживання найжирніших особин є основним натхненням цього алгоритму, де найкращі особини, як правило, беруть більшу участь у покращенні поганих розв'язків. Алгоритм PSO моделює старіння стад птахів або зграй оленів.

Розглянемо далі детальніше Grasshopper Optimization Algorithm (GOA).

Стрибунці - це комахи. Вони вважаються шкідниками через шкоди, яку вони завдають рослинництву та сільському господарству. Хоча стрибунців у природі зазвичай бачать поодиночі, вони об'єднуються в один з найбільших роїв серед усіх істот [14]. Розмір рою може мати континентальні масштаби та бути жахливим для масштабу континенту і є справжнім кошмаром для фермерів. Унікальність рою стрибунців полягає в тому, що ройова поведінка притаманна як у німф, так і в імаго [15]. Мільйони стрибунців-німф стрибають і рухаються, як циліндри, що котяться. На своєму шляху вони поїдають майже всю рослинність. Після такої поведінки, коли вони стають дорослими, вони утворюють рій у повітрі. Так стрибунці мігрують на великі відстані.

Основною характеристикою рою на стадії личинки є повільний рух і маленькі кроки стрибунців. На противагу цьому, великі відстані та різкі рухи на великі відстані є основною рисою рою у дорослому віці. Ще однією важливою характеристикою рою є пошук джерел їжі. Як ми вже обговорювали у вступі, алгоритми, натхненні природою, логічно поділяють процес пошуку на дві тенденції: розвідку та експлуатацію. Під час експлуатації пошукові агенти заохочуються до пошуку, в той час, як під час експлуатації вони, як правило, рухаються локально. Ці дві функції, а також пошук цілей, стрибунці виконують природним чином. Тому, якщо ми знайдемо спосіб математичного моделювання цієї поведінки, ми

зможемо розробити поведінку, ми зможемо розробити новий алгоритм, натхненний природою.

Математична модель, що використовується для моделювання ройової поведінки стрибунців виглядає наступним чином [16]:

$$X_i = S_i + G_i + A_i,$$

де X_i визначає положення i -го стрибунця, S_i – соціальна взаємодія, G_i - сила гравітації на i -го стрибунця, і A_i показує адвекцію вітру.

Потрібно зазначити, що дана модель враховує розташування поточного стрибунця відносно інших стрибунців. Фактично, ми розглянули стан усіх стрибунців, щоб визначити місцеперебування пошукових агентів навколо цілі. Це відрізняється від PSO як найбільш визнаного в літературі методу ройової розвідки. У PSO для кожної частинки є два вектори: позиція і вектор і вектор місцеперебування. Однак у GOA для кожного агента пошуку є лише один вектор положення пошукового агента в GOA. Інша основна відмінність між цими двома алгоритмами полягає в тому, що PSO оновлює положення частинок з повторним поточного положення, особистого та глобального найкращих результатів. Однак, GOA оновлює позицію пошукового агента на основі його поточної на основі його поточної позиції, глобального найкращого результату та позицій всіх інших пошукових агентів. Це означає, що в PSO жодна з інших частинок не робить внеску в оновлення положення частинки, тоді як GOA вимагає, щоб усі пошукові агенти беруть участь у визначенні наступної позиції кожного пошукового агента, тоді як GOA вимагає, щоб усі пошукові агенти брали участь у визначенні наступного положення кожного.

III. Алгоритм навчання нейронної мережі Forward-Forward

Алгоритм Forward-Forward – це жадібна багатошарова процедура навчання, натхненна машинами Больцмана та Noise Contrastive Estimation. Ідея полягає в тому, щоб замінити прямий і зворотний проходи backpropagation двома прямими проходами вперед, які працюють точно так само, як і один, але на різних даних і з протилежними цілями. Позитивний прохід працює з реальними даними і коригує ваги, щоб підвищити якість кожного прихованого шару. Негативний прохід оперує з "негативними даними" і коригує ваги для зменшення якості в кожному прихованому шарі. У цій статті досліджуються дві різні міри якості – сума квадратів нейронних активностей та від'ємна сума квадратів активностей, але існує багато інших мір.

Припустимо, що функція корисності для шару є просто сумою квадратів активностей випрямлених лінійних нейронів у цьому шарі. Мета навчання полягає у тому, щоб зробити функцію доброякісності значно вищою за деяке порогове значення для реальних даних і значно нижчою за це значення для від'ємних даних. Точніше, мета полягає в тому, щоб правильно класифікувати вхідні вектори як позитивні або негативні дані, коли ймовірність того, що вхідний вектор є позитивним (тобто реальним), задається шляхом застосування логістичної функції σ до якості, мінус деякий поріг, θ :

$$p(\text{positive}) = \sigma\left(\sum_j y_j^2 - \theta\right),$$

де y_j - активність прихованої одиниці j до нормалізації шару. Негативні дані можуть бути передбачені нейронною мережею за допомогою низхідних зв'язків або отримані ззовні.

Навчання декількох шарів представлення за допомогою простої пошарової функції доброякісності

Легко побачити, що один прихований шар можна вивчити, якщо зробити так, щоб сума квадратів активностей прихованих одиниць була високою для додатних даних і низькою для від'ємних. Але якщо активність першого прихованого шару використовуються як вхідні дані

для другого прихованого шару, то відрізнити позитивні дані від негативних даних, просто використовуючи довжину вектора активності в першому прихованому шарі. Немає необхідності вивчати будь-які нові ознаки. Щоб запобігти цьому, FF нормалізує довжину прихованого вектора перед використанням його як вхідні дані для наступного шару. Це видаляє всю інформацію, яка була використана для визначення доброти в першому прихованому шарі, і змушує наступний прихований шар використовувати інформацію про відносну активність нейронів у першому прихованому шарі. Ця відносна активність не залежить від нормалізації шарів. Іншими словами, вектор активності у першому прихованому шарі має довжину та орієнтацію. Довжина використовується для визначення доброчесності для цього шару, і лише орієнтація передається наступному шару.

IV.FFNN-оптимізації Grasshopper Optimization Algorithm

В контексті FFNN-оптимізації Grasshopper Optimisation Algorithm застосовано наступним чином:

1. Ініціалізація: Випадковим чином ініціалізуються ваги та зміщення FFNN в межах заданого діапазону. Генерується початкова популяція стрибунців, кожен з яких представляє потенційний розв'язок.

2. Оцінка: Оцінюється придатність кожного стрибунця шляхом навчання FFNN з використанням поточного набору ваг та зміщень. Пристосованість зазвичай визначається продуктивністю мережі на перевіркового наборі даних, де менша помилка або вища точність вказує на кращу пристосованість.

3. Рух стрибунця: GOA використовує різні стратегії руху для дослідження простору пошуку. Ці стратегії включають пошук ділянок з їжею (тобто локальний пошук), слідування за найкращими кониками (тобто глобальний пошук) та стрибки на нові позиції.

4. Локальний пошук: Випадковий вибір деяких стрибунців для локального пошуку шляхом зміни їхніх позицій у межах невеликого району. Це допомагає використовувати локальні області пошукового простору і вдосконалювати рішення.

5. Глобальний пошук: Визначити найкращого стрибунця з найвищим fitness значенням та оновити позиції інших стрибунців, щоб наблизитися до позиції найкращого стрибунця. Це заохочує дослідження перспективних областей у просторі пошуку.

6. Стрибки: Випадковий вибір деяких стрибунців для виконання стрибків шляхом випадкового оновлення їхніх позицій у просторі пошуку. Це дозволяє алгоритму уникати локальних оптимумів і досліджувати нові регіони.

7. Оновити FFNN: Після переміщення стрибунців оновлюємо ваги та зсуви FFNN на основі позицій стрибунців. Цей крок передбачає коригування параметрів за допомогою відповідних математичних операцій, таких як кросовер і мутація, для покращення загальної пристосованості мережі.

8. Завершення роботи: Повторюємо кроки 2-7 до тих пір, поки не буде виконано умову завершення, наприклад, досягнення максимальної кількості ітерацій або бажаного рівня пристосованості. В кінці процесу оптимізації FFNN отримає оптимізовані ваги та зміщення, які можна використовувати для точних прогнозів на невидимих даних.

V. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Серед ключових напрямів подальшого розвитку моделі нейронної мережі Forward-Forward, яка використовує алгоритм оптимізації Grasshopper Optimization Algorithm (GOA) - розширення архітектури мережі, оптимізація параметрів GOA, застосування моделі до реальних даних, порівняння з іншими алгоритмами, розробка додаткових функцій GOA, використання ансамблевих моделей та адаптація до різних завдань. Виконання цих досліджень має потенціал покращити ефективність та універсальність розробленої моделі, сприяючи розвитку області машинного навчання.

VI. ВИСНОВКИ

У даній роботі представлено нейронну мережу типу Forward-Forward із використанням Grasshopper Optimization Algorithm як оптимізатора. Дана нейронна мережа сприяє ефективному розв'язанню задач які ставляться перед нейронними мережами, а також розвитку нейронних мереж та оптимізаторів розглянутих типів та може бути корисною для науковців, аналітиків та інших фахівців, що працюють у даній галузі. Для подальшого розвитку даної роботи у майбутньому, можна застосувати принципи snapshot unsambly та cycle learning rate, що повинно збільшити ефективність нейронної мережі.

ЛІТЕРАТУРА

- [1] C.A. Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art", *Comput Meth Appl Mech Eng*, vol. 191, pp. 1245–87, 2002.
- [2] R.T. Marler, J.S. Arora, "Survey of multi-objective optimization methods for engineering", *Struct Multidiscipl Optim*, vol. 26, pp. 369–95, 2004.
- [3] S. Mirjalili, S.M. Mirjalili, A. Lewis, "Grey wolf optimizer", *Adv Eng Softw*, vol. 69, pp. 46–61, 2014.
- [4] J.C. Spall, "Introduction to stochastic search and optimization: estimation, simulation, and control", John Wiley & Sons, 2005.
- [5] D. Dasgupta, Z. Michalewicz, "Evolutionary algorithms in engineering applications", Springer, 1997.
- [6] X-S. Yang, "Nature-inspired metaheuristic algorithms", Luniver press, 2010.
- [7] S. Mirjalili, A. Lewis, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization", *Swarm Evol Comput*, vol. 9, pp. 1–14, 2009..
- [8] L. Davis, "Bit-climbing, representational bias, and test suite design", *ICGA*, 1991.
- [9] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, "Optimization by simulated annealing", *Science*, vol. 220, pp. 671–80, 1983.
- [10] J.H. Holland, "Genetic algorithms", *Sci Am*, vol. 26, pp. 66–72, 1992.
- [11] R.C. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory", *Proceedings of the sixth international symposium on micro machine and human science*, pp. 39–43, 1995.
- [12] A. Colomi, M. Dorigo, V. Maniezzo, "Distributed optimization by ant colonies", *Proceedings of the first European conference on artificial life*, pp. 134–42, 1991.
- [16] R. Storn, K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces", *Glob Optim*, vol. 11, pp. 341–59, 1997.
- [13] S.J. Simpson, A. McCaffery, "A behavioral analysis of phase change in the desert locust", *Biol Rev*, vol. 74, pp. 461–80, 1999.
- [14] S.M. Rogers, T. Matheson, E. Despland, T. Dodgson, M. Burrows, S.J. Simpson, "Mechanosensory-induced behavioural gregarization in the desert locust *Schistocerca gregaria*", *Exp Biol*, vol. 206, pp. 3991–4002, 2003.
- [15] C.M. Topaz, A.J. Bernoff, S. Logan, W. Toolson, "A model for rolling swarms of locusts", *Eur Phys J Special Top*, vol. 157, pp. 93–109, 2007.

Теорія та практика реалізації базових функцій інтелектуальної АСУ в нейромережевому базисі

Сергій Альошин, Олена Гайтан

*Кафедра комп'ютерних та інформаційних технологій і систем
Національний університет «Полтавська політехніка імені Юрія Кондратюка»
Полтава, Україна*

Анотація—Робота присвячена узагальненню та систематизації процесу синтезу моделей базових функцій управління об'єктами в інтелектуальних автоматизованих системах управління (ІАСУ) для різних предметних галузей у базисі стандартних нейроемуляторів нейронних мереж (Matlab, Statistika та ін.). Такий підхід не вимагає значних витрат, оскільки ці пакети аналізу даних є у відкритому доступі. Синтез нейромоделей у цій ситуації спирається творчий підхід у виборі типу мережі, її складності, методу навчання, функції активації, початкових умов та інших.

Ключові слова—*Інтелектуальна АСУ, Нейронна мережа, Базові функції, Класифікатор, Керуючі впливи.*

I. ВСТУП

У сучасних умовах набирає розвитку технологічний уклад, що ґрунтується на впровадженні у виробництво та управління нових інформаційних технологій, інтелектуальних автоматизованих систем управління. При цьому є необхідність постійного пошуку компромісу у дилемі: «ефективність – вартість». Вихід – знайти оптимальну технологію та інструментарій, що задовольняють обраному критерію «ефективність – вартість».

Напрямок дослідження – синтез моделей базових функцій управління (розпізнавання, прогнозування, адаптація вхідних факторів до цільової функції) на основі відомих оптимальних вирішальних правил (функції правдоподібності, апостеріорної ймовірності, середнього ризику тощо) на базі стандартних пакетів технічного аналізу даних (нейроемуляторів). З цією метою було систематизовано алгоритм синтезу моделей та показано на прикладах реалізованість та ефективність такого підходу.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Питання реалізації базових функцій інтелектуальних систем управління розглянуто в багатьох роботах закордонних та українських дослідників. Проте універсальної програмної бази синтезу моделей базових функцій управління об'єктами для різних сфер діяльності у відкритих публікаціях не проглядається. Для кожної предметної області необхідно синтезувати власні моделі, що враховують потужність вибірки прецедентів, їх інформативність, типи змінних та їх розмірність, взаємну кореляцію, вид розподілу та амплітуду розмаху.

Враховуючи проведений аналіз останніх досліджень і публікацій, було виділено мету дослідження. Мета дослідження – запропонувати методіку та інструментарій моделювання базових процесів об'єктів різної природи у середовищі стандартних емуляторів нейронних мереж із забезпеченням вихідних умов якості та вартості.

III. МЕТОДИ

Під час створення інтелектуальної АСУ послідовно ставляться і вирішуються такі задачі [1,2]:

1.1. Формалізація предметної галузі

По суті, це завдання є завданням кодування. Складається список узагальнених класів, до яких можуть належати конкретні реалізації об'єктів, а також список ознак, якими ці об'єкти в принципі можуть мати.

1.2. Формування репрезентативної навчальної вибірки

Навчальна вибірка є базою даних, що містить описи конкретних реалізацій об'єктів мовою ознак, доповнену інформацією про належність цих об'єктів до певних класів розпізнавання.

1.3. Навчання класифікатора

Навчальна вибірка використовується для формування узагальнених образів класів розпізнавання на основі узагальнення інформації про те, які ознаки мають об'єкти навчальної вибірки, що належать до цього класу та інших класів.

1.4. Оцінка та вибір розмірності простору ознак

Після навчання системи розпізнавання (отримання статистики розподілу частот ознак за класами) стає можливим визначити для кожної ознаки його цінність для розв'язання задачі розпізнавання. Після цього найменш цінні ознаки можуть бути видалені із системи ознак. Потім система розпізнавання повинна бути навчена заново, так як в результаті видалення деяких ознак статистика розподілу ознак, що залишилися, за класами змінюється. Цей процес може повторюватися, тобто бути ітераційним.

1.5. Синтез моделі та демонстрація її адекватності

Реалізується у форматі апроксимації складної функції зв'язку цільових індикаторів з вектором множини вхідних факторів. Після встановлення детермінованого зв'язку факторів та станів об'єкта необхідно забезпечити та обґрунтувати адекватність моделі.

1.6. Технологія знаходження керуючих факторів

Завдання класифікації полягає у тому, що для даного об'єкта за його відомими ознаками системою встановлюється його належність до деякого раніше невідомого класу. Потім нейромережевою моделлю встановлюється, які ознаки найбільш характерні для об'єктів даного класу, а які ні (або які об'єкти навчальної вибірки відносяться до цього класу).

При створенні та експлуатації АСУ ключовою підсистемою вважають автоматизовану систему класифікації та адаптації входів до цільової функції або її конкретних індикаторів. Схема реалізації даних аналітичних функцій представлена та розкрита достатньою мірою у роботах [3,6] і в даному випадку може бути задіяна для практичного запуску базових функцій ІАСУ (рис. 1).

Даний узагальнений підхід до вирішення завдань аналізу вхідних даних, класифікації станів, регресійного аналізу, пошуку керуючих факторів адекватних цільової функції та вхідні дані використовується у широкому спектрі сфер діяльності. Виходячи з заявлених цілей і завдань моделювання базових процесів об'єктів різної природи, використовуються вхідні дані в різних предметних областях і відповідні пакети наборів даних.

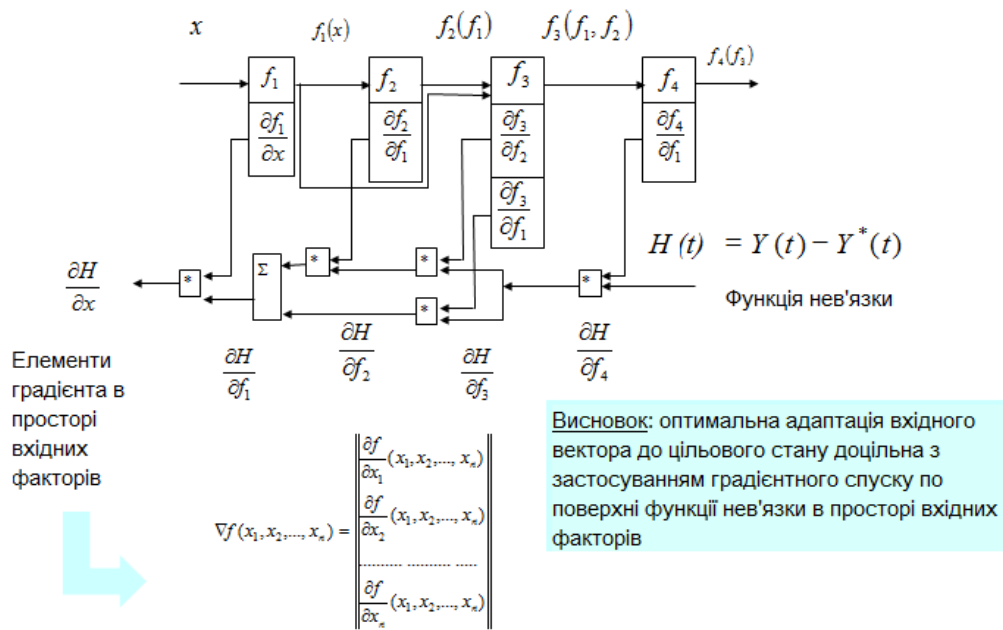


Рисунок 1. Узагальнена схема реалізації аналітичних функцій ІАСУ

Зокрема, для моделювання економічних показників було використано дата сет з економіки та фінансів Quandl (<https://www.quandl.com/>), джерело економічних та фінансових даних при побудові моделей для прогнозування економічних показників або цін на акції, та IMF Data (<https://www.imf.org/en/Data>), що містить дані про міжнародні фінанси, показники боргу, валютні резерви, інвестиції та ціни на сировинні товари.

Обґрунтування та реалізація базових функцій ІАСУ

Інструментально завдання вирішується у базисі існуючих градієнтних методів навчання штучної нейронної мережі алгоритмом зворотного поширення помилки [5-7].

Для моделювання класифікатора доцільно дослідити можливості пакету технічного аналізу даних StatSoft з нейромережним модулем STATISTICA Neural Networks [4,5]. Тоді синтез моделей базових процесів реалізується у просторі процедур, що прискорюють адаптацію моделей у межах обраних обмежень.

На вхід мережі надходить множина пар навчальних векторів $\{x, d\}$, де $\{x\}$ – вхідний вектор, а $\{d\}$ – істинний вихідний вектор ознак об'єкта, що досліджується, $\{y\}$ – множина реакцій нейронної мережі на вхід $\{x\}$. Різниця між $\{y\}$ та $\{d\}$ ($E = \|y - d\|$) – помилка навчання. При середній квадратичній формі міри помилки отримаємо:

$$E = \frac{1}{SM} \sum_{i=1}^S \sum_{j=1}^M (y_j^i - d_j^i)^2, \quad (1)$$

де S – число навчальних пар, M – розмірність вихідного вектора.

Завдання навчання нейронної мережі зводиться до пошуку таких значень вагових коефіцієнтів $w_{i,j}^{(k)}$, щоб помилка навчання E стала меншою за деяке допустиме значення ε : ($E < \varepsilon$) [4-5].

Реалізувати навчання ансамблю моделей дозволяють відомі швидкодіючі спрощені алгоритми навчання [4], для яких висока швидкість збіжності, як ключовий критерій якості, забезпечується при низьких вимогах похибки обчислення градієнта. Наприклад, алгоритм RPROP (resilient back propagation) [4] вирішує завдання, розраховуючи лише знаки градієнтів і не залежить від точності розрахунку величин похідних, а аналізує лише співвідношення знаків прирощень за правилом:

$$\Delta_l^{(i)} = \left\{ \begin{array}{l} \eta^+ \Delta_l^{(i-1)}, \text{ if } \frac{\partial E(w)^{(i-1)}}{\partial w_l} \cdot \frac{\partial E(w)^{(i)}}{\partial w_l} > 0 \\ \eta^- \Delta_l^{(i-1)}, \text{ if } \frac{\partial E(w)^{(i-1)}}{\partial w_l} \cdot \frac{\partial E(w)^{(i)}}{\partial w_l} < 0 \\ \Delta_l^{(i-1)}, \text{ if } \frac{\partial E(w)^{(i-1)}}{\partial w_l} \cdot \frac{\partial E(w)^{(i)}}{\partial w_l} = 0 \end{array} \right. \quad (2)$$

де $0 < \eta^- < 1 < \eta^+$ і визначається емпіричним шляхом. Величина збільшення коригується фіксованим значенням η^+ у тому випадку, коли алгоритм сходиться до мінімуму і похідна не змінює знак. Це прискорює процес на плоских ділянках і уповільнює пошук у разі пропуску локального мінімуму.

Потім визначаються величини змін ваг відповідно до напрямку зменшення градієнта:

$$\Delta w_l^{(i)} = \left\{ \begin{array}{l} \Delta_l^{(i)} \cdot \text{sgn} \left[\frac{\partial E(w)^{(i)}}{\partial w_l} \right], \text{ if } \frac{\partial E(w)^{(i-1)}}{\partial w_l} \cdot \frac{\partial E(w)^{(i)}}{\partial w_l} \geq 0 \\ -\Delta_l^{(i)}, \text{ if } \frac{\partial E(w)^{(i-1)}}{\partial w_l} \cdot \frac{\partial E(w)^{(i)}}{\partial w_l} < 0 \end{array} \right\}, \quad (3)$$

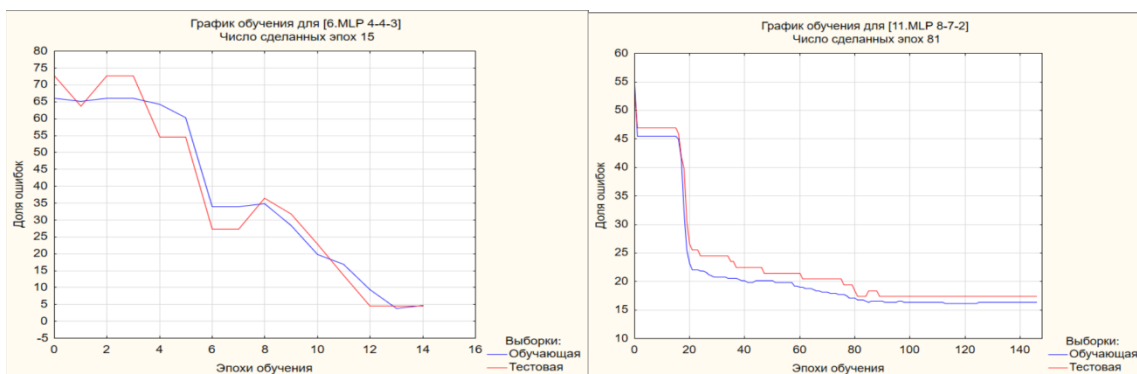
де $\text{sgn}[*]$ – знак функції. Зміна знака похідної помилки під час наступного кроку свідчить про проходження мінімуму функції. Цей результат вимагає повернення до попереднього значення ваги $w_l^{(i-1)}$. Алгоритм ґрунтується на визначенні лише знака добутку похідних функцій на поточному та попередньому кроці. Модифікація синаптичного простору таким способом вимагає значно меншого числа операцій, в порівнянні класичним методом зворотного поширення помилки [4-6].

Аналітичний зв'язок керуючих факторів з цільовою функцією слід встановити, спираючись на теорему Колмогорова-Арнольда про представлення функції кількох аргументів через суму композицій функцій однієї змінної та її адаптації до нейромережевого формату Хехт-Нільсена [4,5]:

$$y(x) = \alpha \sum_{i=1}^H v_i (w_{i1}x_1 + w_{i2}x_2 + \dots + w_{in}x_n + u_i), \quad (4)$$

де H – потужність навчальної вибірки, α, v – параметри нейромережі, n – кількість нейронів, $w_{i1}, w_{i2}, \dots, w_{in}$ – вагові коефіцієнти нейронів.

Демонстрація якості навчання варіантів моделей



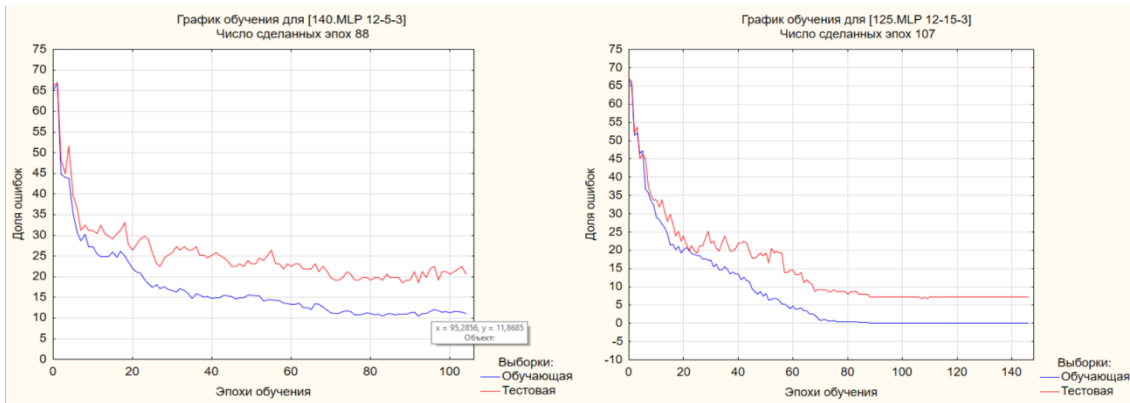


Рисунок 2. Візуалізація динаміки збіжності процесу навчання моделей

Адаптація вхідних факторів до цільової функції

Це завдання вирішується підстроюванням вхідного вектора факторів через градієнт функції помилки за вхідними сигналами мережі [4].

Задавалися індикатори, що відповідають необхідному класу стану об'єкта, та подавалися на вхід навченої мережі. Маючи задану відповідь та відповідь, видану мережею, обчислювався градієнт функції помилки за вхідними сигналами мережі за технологією методу навантажених двоїстих мереж [4,5]. Відповідно до значення елементів градієнта змінювалися значення вхідних сигналів мережі в напрямку зниження помилки, що дозволяє ітераційно отримати вектор вхідних сигналів, що породжують необхідну відповідь.

Таким чином, значення вхідних факторів після закінчення процедури підстроювання є результатом прийняття рішення нейронною мережею щодо переведення системи з одного стану (класу) в інший.

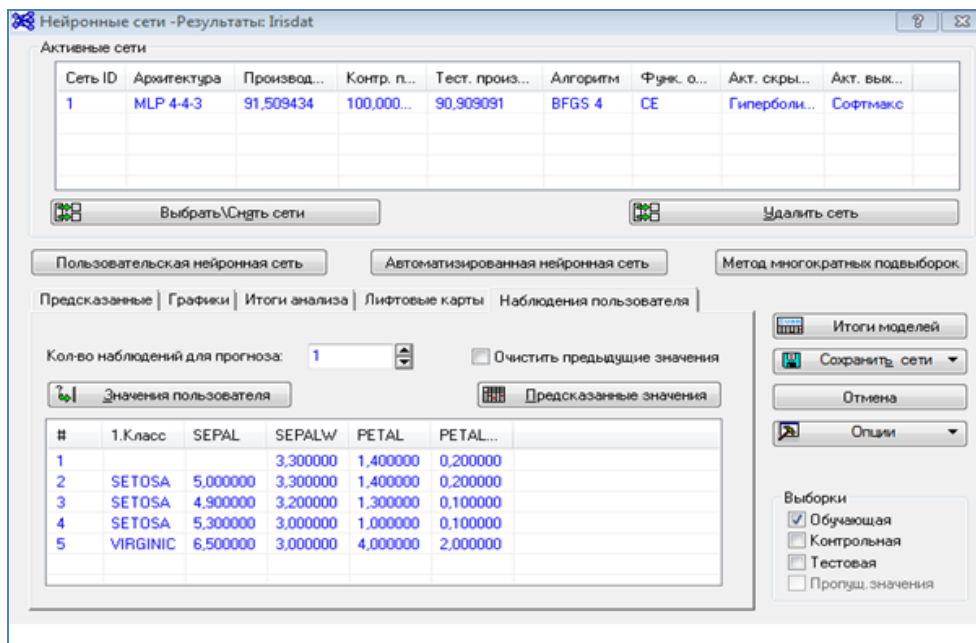


Рисунок 3. Панель пошуку факторів, адекватних цільової функції ІАСУ, та профіль найбільш ефективної моделі

IV. РЕЗУЛЬТАТИ

Для різних умов моделювання на тестових вибірках прикладів отримано ансамбль

продуктивних нейромереж, які якісно підтверджують працездатність запропонованої технології синтезу класифікатора в режимі реального часу. Зберігаємо моделі з найбільш високою продуктивністю:

Сеть ID	Архитектура	Производ...	Контр. п...	Тест. произ...	Алгоритм	Функ. о...	Акт. скры...	Акт. вых...
1	MLP 10-11-1	0.962930	0,000000	0,000000	BFGS 2	Сум. ква...	Тождеств...	Тождеств...
2	MLP 10-4-1	0.977138	0,000000	0,000000	BFGS 3	Сум. ква...	Тождеств...	Тождеств...
3	MLP 10-12-1	0.946570	0,000000	0,000000	BFGS 2	Сум. ква...	Тождеств...	Тождеств...
4	MLP 10-9-1	0.987385	0,000000	0,000000	BFGS 3	Сум. ква...	Тождеств...	Тождеств...
5	MLP 10-4-1	0.982957	0,000000	0,000000	BFGS 2	Сум. ква...	Тождеств...	Тождеств...

Рисунок 4. Профілі моделей із найбільш високою продуктивністю

Адекватність моделювання встановлюється за продуктивністю та помилками на навчальних та тестових множинах, що дозволяє стверджувати про спроможність моделювання базових функцій ІАСУ щодо наборів прецедентів зв'язку факторів та станів досліджуваних об'єктів для широкого спектру предметних областей діяльності [3-6].

Таким чином, реалізовані базові функції АІСУ в середовищі стандартних нейроемулаторів Statistika Neural Network, із заданою продуктивністю і точністю.

V. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Подальші дослідження використання нейромережевого підходу в інтелектуальних автоматизованих системах управління доцільно націлити на часткову або навіть повну автоматизацію базових процесів при вирішенні різних аналітичних завдань у зазначених вище предметних областях, на максимально точні оцінки станів об'єктів аналізу, прогнози динаміки їх змін та знаходження вхідних умов та значень керуючих факторів, що дозволяють змінювати стан (параметри) об'єктів у заданих межах. Повна автоматизація реалізується в рамках наведеної методики аналізу даних та сприяє підвищенню конкурентоспроможності та зниженню витрат від уникнення ручної праці та скорочення часу на прийняття рішення.

VI. ВИСНОВКИ

Нейромережева підтримка реалізації базових функцій інтелектуальної АСУ успішно вирішується із застосуванням класичних пакетів технічного аналізу даних. Синтезовані моделі базових функцій ІАСУ показали прийнятну для практики працездатність. При цьому основні зусилля слід спрямувати на формування бази достатньої інформативної вибірки прецедентів, а синтез моделей здійснювати серед стандартних нейроемулаторів, що істотно скорочує витрати і час при збереженні прийнятної для практики ефективності.

ЛІТЕРАТУРА

- [1] J. C. Giarratano and G. D. Riley, *Expert Systems: Principles and Programming*, 4th Edition, Course Technology, 2004, 288 pp.
- [2] А. А. Морозов та Г. С. Теслер, "Ситуаційне управління та системи підтримки прийняття рішень," в *Збірнику доповідей науково-практичної конференції з міжнародною участю „Системи підтримки прийняття рішень. Теорія і практика”*, Київ, 2005, с. 5–9.
- [3] А. А. Морозов та ін., "Стан та перспективи нейромережевого моделювання СППР у складних соціотехнічних системах," *Математичні машини і системи*, № 1, с. 127–149, 2010.

- [4] S. Haykin, *Neural Networks and Learning Machines*, Prentice Hall, 2009, 906 pp.
- [5] С. П. Альошин та О. М. Гайтан, “Інтелектуальна технологія синтезу автоматизованої експертної системи керування підприємством,” в *Conf. Proc. 2022 International Conference on Innovative Solutions in Software Engineering (ICISSE)*, Ivano-Frankivsk, November 29-30, 2022, pp. 226–231.
- [6] С. П. Альошин, *Нейромережевий базис підтримки рішень у просторі факторів та станів високої розмірності*, Полтава, «Скайтек», 2013, 208 с.

Використання інформаційної системи для формування індивідуальної освітньої траєкторії

Сергій Шаров

*Кафедра комп'ютерних наук
Таврійський державний агротехнологічний університет імені Дмитра Моторного
Запоріжжя, Україна*

Анотація—У статті повідомляється про розробку веб-орієнтованої інформаційної системи для формування індивідуальної освітньої траєкторії здобувачів вищої освіти. З'ясовано, що індивідуальна освітня траєкторія дозволяє забезпечити професійне становлення та саморозвиток здобувача вищої освіти. Зазначається, що використання інформаційних систем в освітньому процесі дозволяє забезпечити підвищення якості управлінської, освітньої, інформаційної діяльності закладу вищої освіти. Описано інструментальні засоби, що були використані для розробки інформаційної системи. У подальших дослідженнях планується дослідити працездатність роботи інформаційної системи в реальних умовах.

Ключові слова—*Індивідуальна освітня траєкторія, Здобувач вищої освіти, Програмне забезпечення, Інформаційна система, Комп'ютерні науки.*

I. ВСТУП

Сучасні реалії вищої освіти актуалізували потребу у розробці та впровадженні ефективних технологій та методологій підготовки майбутніх фахівців, здатних до ефективної комунікації та виконання своїх професійних обов'язків, саморозвитку та самовдосконалення. Особливо це стосується ІТ-фахівців, які повинні постійно слідкувати за технологічним розвитком, бути обізнаними у сучасних цифрових технологіях, пристроях, методах обробки інформації тощо.

Дієвими напрямами організації освітнього процесу є застосування компетентнісного та особистісно-орієнтованого підходів. Можливість формування індивідуальної освітньої траєкторії надає здобувачам вищої освіти можливості розробляти власні освітні програми, які складаються з обов'язкових освітніх компонент, а також дисциплін вільного вибору. Це дозволить сформувати сукупність фахових та загальних компетентностей, які забезпечать успішне працевлаштування та професійну діяльність. Залучення здобувачів освіти до формування власного професійного зростання супроводжується підвищенням мотивації та відповідальності за власні результати навчання.

Стрімкий технологічний та технічний розвиток сприяють формуванню інформаційного суспільства та широкому використанню інформації, що є невичерпним та поновлюваним ресурсом людства разом з енергією та матеріальними ресурсами. Це актуалізація потреби у розробці та оновленні різноманітних програмних продуктів та електронних ресурсів, що дозволяють накопичувати, обробляти, передавати інформацію. До таких програмних засобів відносяться інформаційні системи (ІС), які знайшли своє застосування у різноманітних областях, зокрема освіти, виробництві, економічній діяльності тощо. Водночас, бракує електронних ресурсів, які б дозволяли формувати індивідуальну освітню траєкторію здобувачів вищої освіти, враховуючи освітню програму спеціальності та побажання майбутніх фахівців.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

І. Каньковський, М. Пікуляк, П. Федорук, Л. Немець, К. Сегіда досліджували питання, присвячені аналізу ролі індивідуальних освітніх траєкторій для забезпечення якості вищої освіти. Для реалізації індивідуального підходу та більш якісного навчання використовуються різноманітні програмні засоби та електронні ресурси, зокрема програмно-педагогічні засоби (В. Лапінський). Для побудови індивідуальної освітньої траєкторії учнів використовуються можливості інтегрованого інформаційного ресурсу (Ю. Богачков, В. Милашенко та ін.), SCORM пакети (А. Матвійчук). Реалізація індивідуальної освітньої траєкторії в закладі вищої освіти здійснюється через комплекс сервісів в межах системи електронного навчання (Т. Горбач), застосування програмно-педагогічних засобів (В. Григор'єва).

Можливості інформаційних систем в різних сферах діяльності закладу вищої освіти досліджували Т. Вакалюк, М. Вінник, В. Гриценко, А. Морозов, М. Полторацький. Науковці Н. Задорожна, В. Петрушко, С. Тукало досліджували використання інформаційних систем у науковій діяльності. Водночас, не зважаючи на різноманіття електронних ресурсів для забезпечення освітньої діяльності, бракує досліджень, присвячених розробці та використанню інформаційних систем для формування індивідуальної освітньої траєкторії у вищій школі.

Враховуючи зазначені вище тенденції, метою нашої статті є повідомлення про розробку інформаційної системи для формування індивідуальної освітньої траєкторії здобувача вищої освіти.

III. МЕТОДИ

Для реалізації поставленої мети нами були використані теоретичні та емпіричні методи. Зокрема, аналіз та синтез наукової та спеціальної літератури з інформатики здійснювався з метою визначення сутності індивідуальної освітньої траєкторії, призначення та особливостей інформаційних систем. Для визначення вибіркового освітніх компонентів було здійснено аналіз освітньої програми та навчальних планів підготовки здобувачів вищої освіти спеціальності 122 «Комп'ютерні науки». Також були використані емпіричні методи: опитування здобувачів освіти щодо розуміння індивідуальної освітньої траєкторії; порівняння можливостей інструментальних засобів для розробки веб-орієнтованої інформаційної системи для формування індивідуальної освітньої траєкторії.

У відповідності до етапів життєвого циклу інформаційної системи, нами були проведені наступні види робіт: аналіз освітньої програми «Комп'ютерні науки», за якою здійснюється підготовка здобувачів вищої освіти у ТДАТУ імені Дмитра Моторного; визначення основних модулів інформаційної системи, що будуть забезпечувати реалізацію основної мети даної ІС; проектування та розробка інформаційної системи для формування індивідуальної освітньої траєкторії; часткове наповнення системи інформацією про освітні компоненти, викладачів, здобувачів освіти; перевірка працездатності розробленої системи на локальному сервері.

IV. РЕЗУЛЬТАТИ

Науковці наголошують на важливості формування індивідуальної освітньої траєкторії здобувачів вищої освіти, оскільки це надасть можливість забезпечити їх загальнокультурний, професійний, особистісний розвиток та саморозвиток. Реалізація індивідуальної освітньої траєкторії дозволить стимулювати у здобувачів вищої освіти більшу самостійність до активної участі у процесі проектування власної освітньої діяльності.

За визначенням, під індивідуальною освітньою траєкторією розуміється персональний шлях здобувача освіти до реалізації його особистісного потенціалу, що ґрунтується на виборі форм, методів та засобів навчання, освітніх програм та навчальних дисциплін, що входять до їх складу [1]; обраний за власним бажанням рух до досягнення рівня професійної компетентності у відповідності до стандарту освіти при постійній педагогічній, професійній підтримці та контролі [2]; персональний рух здобувача освіти до особистісно-професійного зростання, що характеризується самостійністю, активністю та відповідальністю за свій вибір.

На думку В. Лапінського, наразі спостерігається ситуація, яку можна охарактеризувати як зародження нового освітнього середовища, необхідними складовими якого на всіх рівнях стали інформаційно-комунікаційні технології (ІКТ) [3]. Їх упровадження у процес опанування освітніми компонентами відкриває вагомі перспективи покращення теоретичної бази знань та практичної спрямованості навчання, розкриття креативного потенціалу здобувачів освіти та викладачів у відповідності до їх індивідуальних запитів, нахилів та здібностей. Водночас, якщо раніше поєднання традиційних форм і методів навчання з комп'ютерними технологіями відбувалося на рівні забезпечення контролю знань у вигляді комп'ютерного тестування або подання навчального матеріалу в електронному вигляді, зараз звертаємо увагу на значне розширення можливостей ІКТ щодо організації освітнього процесу та забезпечення процесу опанування знаннями.

Якщо брати до уваги інформаційні системи, то їх використання в освіті та науці дозволяє автоматизувати управлінську, освітню та інформаційну діяльність закладу освіти, здійснити дієві кроки на шляху забезпечення цифровізації освіти. Як зазначається у роботі [4], інформаційні системи успішно впроваджуються на рівні закладу вищої освіти. Вони мають багатомодульну структуру та в залежності від масштабу можуть складатися з управлінського, організаційного, освітнього та наукового модулів. Розвиток штучного інтелекту вплинув на появу сучасних інформаційно-аналітичних систем управління закладом вищої освіти [5]. Автором була успішно впроваджена інформаційна система у процес проходження навчально-виробничих практик. Вона забезпечувала оформлення звітності з виробничої практики в електронному форматі [6]. Інший тип ІС можна використовувати для формування рейтингу здобувачів вищої освіти за різними критеріями, зокрема результатами контрольних заходів [7]. В контексті наукової діяльності отримала визнання з боку науковців інформаційна система менеджменту наукових досліджень [8]. Все це свідчить про різноманітність сфер застосування сучасних ІС та їх актуальність. Слід зазначити, що всі вище перераховані інформаційні системи є веб-орієнтованими, що зараз в умовах дистанційного навчання робить їх доволі актуальними та корисними.

Враховуючи широкі можливості інформаційно-комунікаційних технологій з метою прискорення процесу формування індивідуального навчального плану спеціальності, нами була розроблена інформаційна система для формування індивідуальної освітньої траєкторії здобувачів другого (магістерського) рівня вищої освіти. Головне завдання електронного ресурсу полягає у забезпеченні вибору здобувачами освіти дисциплін, які вони будуть вивчати у межах освітньої програми спеціальності 122 «Комп'ютерні науки». Також в системі реалізована можливість створення та редагування навчального плану, відображення дисциплін, що входять до навчального плану, анотацій до них (у процесі подальшої розробки анотації передбачається замінити силабусами) та викладачів, що закріплені за цими дисциплінами, відображення дисциплін, обраних зареєстрованим здобувачем освіти, формування звітів на основі збережених даних. Його використання повинно вплинути на якісні показники навчального процесу, зокрема на підвищення успішності з окремих дисциплін вільного вибору.

Загальний принцип роботи з інформаційною системою наступний: адміністратор (представник деканату, завідувач кафедри, гарант) створює навчальні плани з переліком дисциплін вільного вибору студентів, призначає їм викладачів. Після цього відбувається реєстрація здобувачів вищої освіти в системі, після чого вони обирають необхідну кількість дисциплін вільного вибору з урахуванням загальної кількості кредитів. Після заповнення здобувачами освіти власної освітньої траєкторії адміністратор має можливість побачити обрані дисципліни, кількість здобувачів освіти, які планують її вивчати, викладачів, які отримують навантаження тощо.

Оскільки технічним завданням передбачалось, що інформаційна система буде працювати у віддаленому та багатокористувацькому режимах, у якості інструментального засобу для розробки web-ресурсу була використана безкоштовна мова програмування PHP. Дані здобувачів освіти, а також службова інформація інформаційної системи зберігається у вільному програмному забезпеченні MySQL. Для розробки структури бази даних було використано безкоштовний програмний продукт PhpMyAdmin. Для забезпечення захисту персональної інформації та виключення можливості інтеграції SQL-ін'єкцій в роботу інформаційної системи, була використана бібліотека PDO (PHP Data Objects). Більш детальний опис функціональних можливостей розробленої інформаційної системи поданий у роботі [9].

V. ОБГОВОРЕННЯ

На нашу думку, розроблена інформаційна система дозволяє вирішити декілька задач в різних вимірах. З одного боку, за допомогою інформаційної системи здобувач вищої освіти створює власну освітню траєкторію, бачить, які освітні компоненти йому необхідні для розвитку професійних та ключових компетентностей. Для завідувача кафедри інформаційна система теж доволі корисна, адже дозволяє створювати навантаження по кафедрі на основі обраних здобувачами освіти вибіркового дисциплін [9]. Для гаранта освітньої програми дана інформаційна система корисна тим, що про неї можна буде написати в звіті самоаналізу при акредитації освітньої програми. Для закладу вищої освіти, за умови внесення даних за різними освітніми програмами, забезпечується процес цифровізації управлінської діяльності.

Слід зазначити, що реалізація індивідуальної освітньої траєкторії зазвичай здійснюється у двох напрямках. Перший напрямок передбачає індивідуалізацію освітнього процесу в межах вивчення окремих дисциплін. В даному випадку здобувач вищої освіти в межах академічної свободи може обирати способи вирішення навчальних завдань, здійснювати самооцінювання, брати участь у неформальній освіті, брати участь в колективній роботі, тренуючи таким чином м'які навички та ін. Інший напрямок полягає у побудові власного освітнього шляху через вибір переліку вибіркових освітніх компонент (дисциплін вільного вибору здобувачів вищої освіти). Таким чином, здобувач освіти буде розвивати у собі саме ті компетентності, які найбільш відповідають його можливостям та нахилам. Звісно, на вибір здобувачем освіти конкретної дисципліни, крім його бажання, впливають також інші фактори, зокрема якість викладання дисципліни, професіоналізм викладача, змістове наповнення курсу тощо. Водночас, найбільш ефективним варіантом побудови власного освітнього та кар'єрного шляху є одночасне використання двох вище зазначених напрямків.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Зараз автори знаходяться на завершальному етапі створення методичних вказівок до використання інформаційної системи, після чого передбачається її імплементація в освітній процес. Тому у подальших дослідженнях планується дослідити працездатність роботи інформаційної системи в реальних умовах.

VII. ВИСНОВКИ

Отже, використання інформаційних систем в освітньому процесі та науковій діяльності дозволяє забезпечити підвищення якості управлінської, освітньої, інформаційної діяльності закладу вищої освіти. Використання інформаційної системи для формування індивідуальної освітньої траєкторії буде на користь безпосередньо здобувачу освіти, завідувачу кафедри, адміністрації університету. Використання мови програмування PHP, бази даних MySQL, програмного забезпечення PhpMyAdmin, бібліотеки PDO та інших інструментальних засобів дозволяло забезпечити якісну роботу інформаційної системи в онлайн режимі.

ЛІТЕРАТУРА

- [1] L. M. Niemets, K. Yu. Sehida, and M. O. Lohvynova, "Individual educational trajectories: role in ensuring the quality of higher education," *Problems of modern education*, vol. 10, pp. 51-60, 2019 [in Ukrainian]
- [2] I. Ye. Kankovskyi, "Individual educational trajectories as a necessity of the modern professional training process of a specialist," *Professional education: problems and prospects*, vol. 4, pp. 62-65, 2013 [in Ukrainian]
- [3] V. Lapinskyi, "Electronic teaching aids - retrospect and prospects," *Informatics*, vol. 33, pp. 3-9, 2011 [in Ukrainian]
- [4] A. V. Morozov and T. A. Vakaliuk, "An electronic environment of higher education institution (on the example of Zhytomyr Polytechnic State University)," *J. Phys. Conf. Ser.*, vol. 1840, no. 1, p. 012061, Mar. 2021, doi: 10.1088/1742-6596/1840/1/012061.
- [5] V. Hrytsenko, *Organizational and pedagogical principles of the creation and implementation of a web-oriented information and analytical system of university management: Monograph*, Cherkasy, Ukraine: Bohdan Khmelnytsky National University of Cherkasy, 2016 [in Ukrainian]
- [6] S. Sharov, and I. Filipov, "The development of the information system of educational and industrial practices," *Physical and mathematical education*, vol. 3(13), pp. 194-198, 2017 [in Ukrainian]
- [7] M. Vinnyk, M. Poltorackiy, Y. Spivakovska, T. Vinnyk, L. Bondarenko, and Y. Revenko, "Measuring the effectiveness of the implementation of individual educational trajectories by university students," in *CEUR Workshop Proceedings*, 2020, vol. 2732, pp. 1286-1297.
- [8] N. Zadorozhna, V. Petrushko, and S. Tukalo, "The information system of scientific research management in the National Academy of Sciences of Ukraine," *Information technologies in education*, vol. 15, pp. 129-137, 2013 [in Ukrainian]
- [9] S. Sharov, D. Lubko, O. Lomeiko, and H. Chemerys, "Information system for the formation of students' individual educational trajectory", 1st International Conference on Education, Humanities, Health and Agriculture, 2021, doi: 10.4108/eai.3-6-2021.2311038.

Набуття soft skills IT фахівцем в процесі ОНЛАЙН НАВЧАННЯ

Ольга Чуб¹, Марина Новожилова²

¹кафедра теоретичної та прикладної системотехніки
Харківський національний університет імені В.Н.Каразіна
Харків, Україна

²кафедра комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О.М.Бекетова
Харків, Україна

Анотація—Роботу присвячено визначенню особливостей та основних вимог до структури хмарного інформаційного освітнього середовища, що забезпечує розвиток навичок командної роботи студентів. Проводиться аналіз базового інструментарію корпоративної інформаційної системи університету та засобів хмарної технології Google Colab, що уможливають розвиток навичок командної роботи як під керівництвом викладача, так і в окремою командою студентів в рамках самостійної роботи.

Ключові слова— *Хмарні технології, Онлайн освіта, М'які компетентності, Робота в команді.*

I. ВСТУП

Під дією зовнішніх факторів непереборної глобальної сили, що безпосередньо вплинули на 70% учнів та студентів у всьому світі [1], освіта в цілому, і IT освіта зокрема, стрімко набувають як онлайн, так і дистанційну форму.

І якщо попит ринку праці на так звані hard skills – технічні навички, професійні знання залишається перманентно високим, змінюючись в деталях та розповсюджуючись на суміжні предметні галузі, то роль гнучких навичок, «м'яких компетентностей», тобто soft skills (англ.) – значно підвищилась.

На відміну від професійних компетентностей (hard skills), які можна чітко виміряти, і які залишаються незмінними незалежно від компанії-роботодавця, soft skills є універсальними компетентностями, які набагато важче виміряти кількісними показниками, і які більше залежать від ситуації, від оточення, корпоративної культури.

Однією з таких м'яких компетентностей, можливо, найважливішою і узагальнюючою, є вміння працювати в команді.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Робота в команді – цю м'яку компетентність для майбутнього програміста, фахівця з розроблення комп'ютерних програм важко переоцінити. Тому при підготовці Стандартів вищої освіти України для чотирьох базових спеціальностей галузі 12 Інформаційні технології: 121 Інженерія програмного забезпечення [2], 122 Комп'ютерні науки [3], 123 Комп'ютерна інженерія [4], 124 Системний аналіз [5], а також для традиційно суміжної спеціальності 174 Автоматизація, комп'ютерно-інтегровані технології та робототехніка визначено, що однією з обов'язкових загальних компетентностей, обумовлених цими Стандартами, є здатність працювати в команді. Це є відбиттям загальноосвітнього досвіду складання відповідних нормативних документів та регламентів.

Зазначимо, що вміння працювати в команді є одним із вимірів узагальненої чотирирівневої моделі навчання Центру редизайну навчальних планів (Center for Curriculum

Redesign (CCR) in Boston, MA) [6, 7], яка забезпечує організаційні рамки компетентностей, необхідних в 21-му сторіччі, і включає наступне:

- Знання, як базові так і професійні, традиційні і нові предмети, міждисциплінарність.
- Навички. Креативність і критичне мислення, уміння працювати в команді (колаборативність) і уміння спілкуватися (комунікабельність).
- Характер: допитливість і наполегливість, ініціативність і саморегуляція, гнучкість і адаптивність, лідерство і відповідальність, громадянськість і культурна свідомість.
- Метанавчання. Здатність до навчання або те, як люди розуміють і адаптуються до навчання. Вона базується на метакогнітивній активності, здатності стратегічно мислити, планувати, контролювати та оцінювати особистий прогрес, а також бажання вчитися.

Стаття [8] присвячена теоретичному аналізу основних можливостей цифрових освітніх сервісів, цифрових платформ і ресурсів в умовах розвитку технологій онлайн навчання.

В публікації [9] проведено аналіз необхідних компетенцій для ІТ професіоналів в галузі науки про дані (Data science), яка найбільш показовою, міждисциплінарною, що об'єднує такі галузі, як інформатика, аналіз даних і статистика, і поєднує їх з іншими спеціалізаціями, які можуть включати функції корпоративної підтримки, такі як структурування, аналіз та управління даними, візуалізацію даних, управління базами даних та аналіз процесів тощо.

Повертаючись до сучасних умов навчання та виконання професійних завдань в цілому, необхідно відзначити, що збереження неперервності освіти за умов пандемії COVID-19 та повномасштабної війни в Україні став можливим тільки завдяки прогресу в галузі хмарних технологій.

II. МЕТОДИ

Досвід авторів викладання як базових, так професійно-орієнтованих дисциплін спеціальностей галузі знань 12 Інформаційні технології в умовах забезпечення онлайн освіти дозволив дослідним шляхом сформулювати основні вимоги до структури хмарного інформаційного освітнього середовища. Такими вимогами є:

- гнучкість у використанні,
- забезпечення неперервності навчання,
- мінімум програмних засобів, що використовуються,
- мінімум обчислювально-апаратних ресурсів учасників освітнього процесу,
- захист персональної інформації.

Виконання саме цих умов дозволяє забезпечити ефективність освітнього процесу, який проводиться в онлайн форматі.

Сьогодні спектр засобів програмно-апаратного забезпечення досить широкий, що уможливує різні конфігурації освітнього середовища. Керуючись досвідом використання різних налаштувань, можна запропонувати наступну структуру хмарного інформаційного освітнього середовища (рис. 1), що містить корпоративну систему закладу вищої освіти, середовище MS TEAMS, систему Moodle, які в цілому складають струнку систему підтримки освітнього процесу.

Стороннім, але дуже важливим елементом цієї системи автори вважають застосування програмного середовища Google Colaboratory [10] як базового засобу забезпечення інтерактивності навчання, організації роботи в команді, розвитку креативності мислення, комунікаційних компетентностей, набуття навичок розподілу задач та обов'язків при створенні програмного продукту за короткий визначений час.

III. РЕЗУЛЬТАТИ

Google Colaboratory або Colab [10] – це інтегроване середовище розробки, що містить редактор початкового коду, а також інструменти для автоматизації складання та тестування

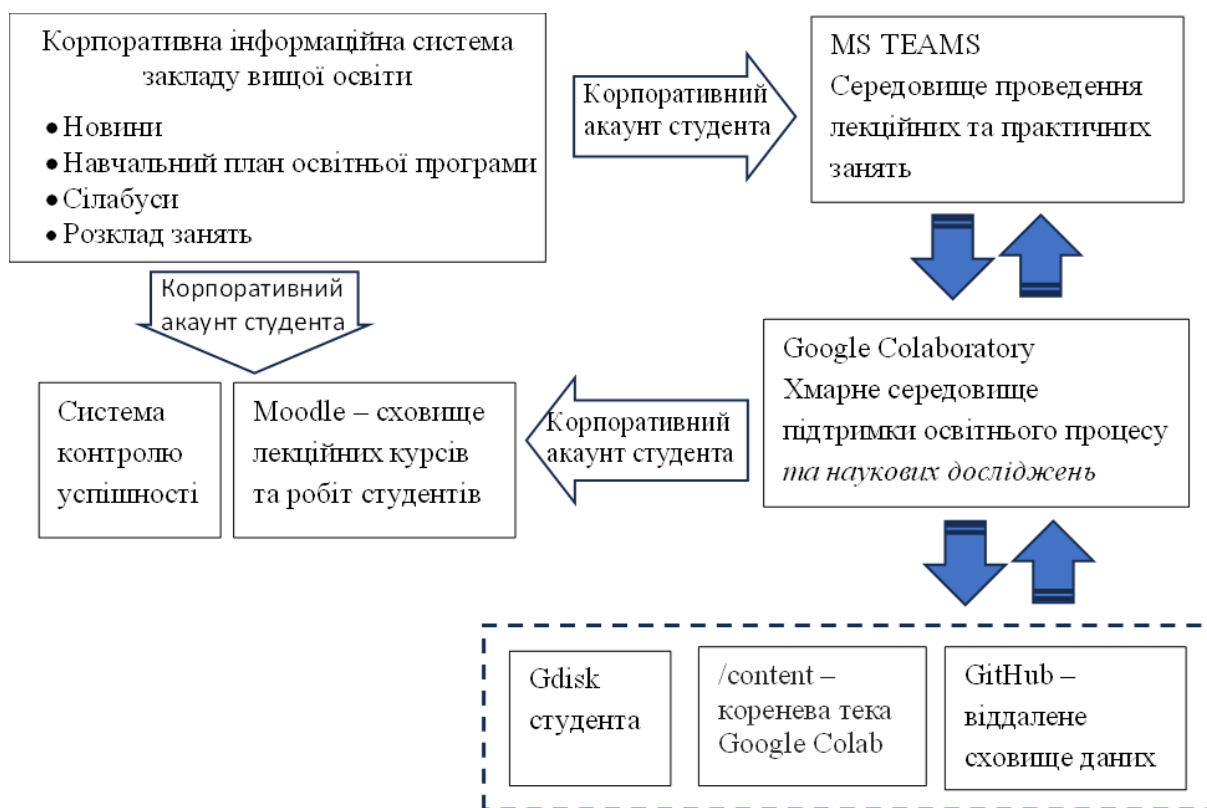


Рис. 1 – Інфологічна модель хмарного інформаційного освітнього середовища

програм. Це середовище є безкоштовним хмарним сервісом від Google Research, заснованим на Jupyter Notebook (<http://www.jupyter.org>).

Середовище Colab дозволяє налаштувати сумісну командну роботу для групи студентів, а також сумісну роботу студентів та викладача над одним файлом або однією задачею. Дружній інтерфейс Colab дозволяє налаштування сумісної роботи над файлами з іншими користувачами, створювати коментарі, корегувати код і в цілому робити все, що можна виконувати у Google Документах. Тому при загальному доступі до блокноту Colab весь його вміст буде доступним іншим членам команди (текст, код, коментарі, вихідні дані).

IV. ОБГОВОРЕННЯ

Google Colaboratory як хмарний додаток, вимагає мінімум апаратних ресурсів. Ця програма працює в інтернет-браузерах як на настільних ПК, так і на мобільних пристроях. Ознакою нашого часу є застосування студентами своїх мобільних пристроїв під час навчання, тому середовище Colab незамінне у цьому випадку, дозволяючи студентам керувати навчанням за допомогою свого мобільного телефону. Google Colab доступ до безкоштовного хмарного сховища, що дозволяє працювати без використання зовнішніх пристроїв зберігання.

Проведення інструментального засобу Google Colab як складової інфологічної моделі хмарного інформаційного освітнього середовища в галузі інформаційних технологій дозволяє також організацію та документування безпосереднього зворотнього зв'язку як між студентами та викладачем, так і у групі студентів. Такий зворотній зв'язок дає можливість викладачу корегувати план навчання, а студентам – розвивати та реалізувати ще одну важливу м'яку навичку – вміння комунікувати з іншими членами команди та знаходити спільні рішення на всіх етапах створення програмного забезпечення.

V. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Запропонована структура інформаційного середовища є досить гнучкою. В подальшому передбачається налаштування використання цієї інформаційної структури як платформи для навчання студентів основам проведення наукових досліджень, зокрема у галузі інтелектуального аналізу даних з різних предметних областей діяльності людини, машинного навчання, паралельних обчислень. Побудові такої інноваційної інформаційної технології сприятиме можливість підключення безкоштовних потужних графічних процесорів GPU (забезпечує паралелізацію виконання завдань) та TPU (тензорний процесор, розробка Google, призначений для тренування нейромереж). Крім того, Google Colaboratory дозволяє використовувати Seedbank – репозиторій прикладів інтерактивного машинного навчання, які можна запускати у своєму браузері без налаштувань як базу для розширення та створення власних проєктів та ідей студентів, що є технічною та інформаційною базою розвитку креативного мислення студентів.

VI. ВИСНОВКИ

Створено та протестовано інфологічну модель хмарного інформаційного освітнього середовища, яка відповідає визначеним вимогам та дозволяє виробити у студентів та викладачів дух корпоративної культури. При цьому найбільш ефективним є синхронне навчання, коли взаємодія між викладачем і студентом, а також студентами між собою (групове навчання) відбувається «тут і зараз», як в офлайн режимі.

Застосування хмарної технології Google Colaboratory дозволяє отримати очевидний синергетичний ефект, тобто ефект від спільної роботи команди якісно перевершує ефект від роботи окремих людей. Отже, робота в команді, спільна робота студентів може в сумі дати набагато більше, ніж результати їх роботи поодиноці.

ЛІТЕРАТУРА

- [1] I. Bouchrika *Online Education Statistics: 2023 Data on Higher Learning & Corporate Training* (2023). Accessed: Oct. 16, 2023. [Online]. Available: <https://research.com/education/online-education-statistics>.
- [2] *Стандарт вищої освіти України 121 Інженерія програмного забезпечення, перший (бакалаврський) рівень*, Жовтень 2018. [Електронний ресурс]. Доступно: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/12/21/121-inzhener.programn.zabezp.bakalavr-1.pdf>
- [3] *Стандарт вищої освіти України 122 Комп'ютерні науки, перший (бакалаврський) рівень*, Липень 2019. [Електронний ресурс]. Доступно: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/2019/07/12/122-kompyut.nauk.bakalavr-1.pdf>
- [4] *Стандарт вищої освіти України 123 Комп'ютерна інженерія, перший (бакалаврський) рівень*, Листопад 2018. [Електронний ресурс]. Доступно: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/123-kompyuterna-inzheneriya.pdf>.
- [5] *Стандарт вищої освіти України 124 Системний аналіз, перший (бакалаврський) рівень*. , Листопад 2018. [Електронний ресурс]. Доступно: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/12/21/124-sistemn.analiz-bakalavr-1.pdf>.
- [6] C. Fadel, B. Trilling, M. Bialik *Four-Dimensional Education: The Competencies Learners Need to Succeed* (2023). Accessed: Oct. 2023. [Online]. Available: <https://curriculumredesign.org/our-work/four-dimensional-21st-century-education-learning-competencies-future-2030>
- [7] N. Akimov, N. Kurmanov, A. Uskelenova, N. Aidargaliyeva, D. Mukhiyayeva, S. Rakhimova, B. Raimbekov and Z. Utegenova «Components of education 4.0 in open innovation competence frameworks: Systematic review». *Journal of Open Innovation: Technology, Market, and Complexity*, vol.9, no 2, 100037, Mar. 2023, doi:org/10.1016/j.joitmc.2023.
- [8] О. С. Пономарьов «Реалізація можливостей цифрових освітніх сервісів і ресурсів в умовах розвитку технологій змішаного навчання», *Педагогічні науки: теорія та практика*, № 2, с. 136-140, 2023. doi.org/10.26661/2786-5622-2023-2-19.
- [9] S. McGrath, Sh.Yamada «Skills for development and vocational education and training: Current and emergent trends». *International Journal of Educational Development*. vol. 102, 102853. Oct. 2023. doi.org/10.1016/j.ijedudev.2023.102853
- [10] *Google Colaboratory*. (2020). Google. Accessed: Oct. 25, 2023 [Online]. Available: <https://colab.research.google.com>.
2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

Порівняння ефективності застосування технологій штрихового кодування та RFID у логістичних процесах

Ігор Невлюдов, Андрій Слюсар, Софія Хрустальова, Кирило Хрустальов, Віктор Косенко

*Кафедра комп'ютерно-інтегрованих технологій, автоматизації та робототехніки
Харківський національний університет радіоелектроніки
Харків, Україна*

Анотація—У сучасному світі логістичні процеси стають все більше автоматизованими, і технологія грає важливу роль у цьому контексті. Ця стаття присвячена порівнянню ефективності застосування двох основних технологій ідентифікації товарів—штрихового кодування і RFID—у логістичних процесах.

Ключові слова—*Логістичні процеси, Ідентифікація, Штрихове кодування, Технологія радіочастотної ідентифікації, Виробнича логістика.*

I. ВСТУП

Логістичні процеси стали невід'ємною складовою сучасного виробництва, які здебільшого спрямовані на оптимізацію та автоматизацію з метою забезпечення найвищого рівня продуктивності і виробничої якості. Постійні зміни в бізнес-середовищі, посилені вимоги до швидкості і точності виробництва, а також потреби у ефективному управлінні запасами та логістиці ставлять перед підприємствами низку складних завдань. В цьому контексті вибір технологій ідентифікації товарів стає критичним завданням для досягнення конкурентної переваги. Одним з ключових аспектів управління логістичними процесами є ідентифікація та відстеження товарів на різних етапах виробництва та постачання. Ця операція вимагає точності, надійності та ефективності, оскільки помилки в ідентифікації або відстеженні товарів можуть призвести до серйозних втрат та помилок в управлінні. Два основних методи ідентифікації, які застосовуються у виробництві та логістиці, – це штрихове кодування та RFID (Radio-Frequency Identification). Штрихове кодування вже давно використовується в різних сферах, і воно відоме своєю простотою, доступністю та низькими витратами на обладнання та інтеграцію. З іншого боку, технологія RFID надає можливість безконтактно ідентифікувати товари та збирати дані в режимі реального часу, але вимагає спеціалізованого обладнання та програмного забезпечення. Для багатьох підприємств виникає питання: яка з цих технологій є більш ефективною в конкретних умовах виробництва та логістики?

Метою даної роботи є проведення порівняльного аналізу ефективності застосування технології штрихового кодування та технології RFID в логістичних виробничих процесах з метою визначення їхніх переваг та недоліків у різних сценаріях виробництва та управління ланцюгом постачання. Ця робота розглядає питання витрат, швидкості ідентифікації, надійності, можливості інтеграції та інших аспектів, які впливають на вибір між цими двома технологіями. Встановити, яка з цих технологій є більш вигідною в конкретних умовах виробництва та логістики, може допомогти підприємствам підвищити ефективність виробництва, зменшити витрати та покращити управління ланцюгом постачання, що стає ключовими завданнями в сучасному бізнес-середовищі. Гіпотеза дослідження полягає в

тому, що RFID може бути більш ефективною технологією ідентифікації та відстеження товарів у порівнянні з штриховим кодуванням.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Розглядаючи останні дослідження та публікації з області порівняння ефективності технології штрихового кодування і технології RFID в логістичних виробничих процесах, можна визначити кілька ключових тенденцій та висновків.

Перший важливий аспект полягає в тому, що обидві технології мають широкий спектр застосувань і показують свою ефективність у різних галузях виробництва та логістики. Однак сучасні дослідження вказують на те, що RFID досить часто переважає у вимірах продуктивності та точності ідентифікації в порівнянні з штриховим кодуванням. Це особливо актуально у сферах, де вимагається швидке та точне відстеження великих обсягів товарів або де потрібна реальна часова інформація про рух товарів [1].

Другий аспект стосується витрат на впровадження і обслуговування обох технологій. Дослідники наголошують, що штрихове кодування має значно менші витрати на інфраструктуру та обладнання. Також воно вже широко використовується у багатьох галузях і має меншу початкову бар'єрну вартість для впровадження. У той час як RFID може потребувати значних інвестицій у спеціалізоване обладнання та програмне забезпечення.

Третій аспект, який варто враховувати, полягає в технічних обмеженнях обох технологій. Штриховий код вимагає прямого видимого контакту із сканером, тому в ньому можуть виникати проблеми з читанням при зношуванні або пошкодженні. З іншого боку, RFID може виявляти обмеження в роботі в металевих або вологих середовищах, а також при високих швидкостях руху товарів [2].

Сучасні дослідження також відзначають тенденцію до поєднання обох технологій в комплексних системах ідентифікації та відстеження. Це дозволяє використовувати переваги обох технологій і зменшує ризики залежності від однієї з них. Однак такі рішення можуть бути складнішими в інтеграції та управлінні. Загалом, останні дослідження та публікації підтверджують, що вибір між технологією штрихового кодування та RFID повинен базуватися на конкретних вимогах та умовах виробництва. Прийняття рішення вимагає зваження переваг та недоліків кожної технології в контексті конкретної ситуації [3].

III. МЕТОДИ

У теперішній час на підприємствах торгівлі застосовується технологія штрихового кодування. Це найбільш використовувана технологія автоматичної ідентифікації, яка з'явилася понад 30 років тому. У ній для ідентифікації фізичних об'єктів застосовуються світлові хвилі. Звичайний штрих-код – двійковий код, який відображається у вигляді впорядкованих паралельних ліній, розділених пробілами. Дана структура являє собою набір цифр або знаків, при цьому смуги і пробіли мають різну ширину [4]. У процесі сканування зчитувач вимірює силу світла, що відбивається від чорних і білих ділянок штрих-коду. Темний штрих поглинає світло, а світлий пробіл його відображає. Фотодіод перетворює світло такої структури в електричний струм (або аналоговий сигнал). Потім електрична схема декодує вироблений струм у цифрові дані. Ці дані і є тим, що було спочатку закодовано штрих-кодом. Найпоширенішою серед систем кодування з використанням штрих-кодів є лінійна система кодування EAN (European Article Number), що з'явилася в 1976 р. і була призначена для торгівлі продовольчими товарами. Штрих-коди недорогі у виробництві, точно відображають інформацію.

Ця технологія дає змогу досягти наступних переваг при переході на безпаперовий документообіг: від 10% до 20% збільшується пропускна спроможність складу; середній час стандартних операцій (приймання, відвантаження, переміщення) у 3 рази; на 40 % знижується ресурс складського запасу; на чверть зменшується кількість складського персоналу; паперовий документообіг скорочується до мінімуму. Але навіть усі перелічені переваги не можуть компенсувати такі недоліки, як невеликий обсяг збереженої інформації та відсутність можливості запису нових даних. Крім цього, технологія штрихового кодування не дозволяє відстежувати товар у русі по всьому логістичному ланцюжку [5]. Тому були зроблені кроки для пошуку нових технологій, які б усунули недоліки, притаманні системам штрихового кодування.

Одним із таких шляхів вирішення є технологія RFID (Radio Frequency Identification) – технологія радіочастотної ідентифікації. Вона забезпечує інформаційну прозорість на всіх ділянках ланцюжка створення вартості від постачальників до кінцевого споживача. Базова RFID-система працює наступним чином. Радіотехнічний пристрій, мітка, прикріплюється до об'єкта, який необхідно ідентифікувати. Унікальні ідентифікаційні дані про об'єкт зберігаються в мітці. Коли позначений об'єкт підноситься до зчитувального пристрою (рідера), мітка передає ці дані в рідер через антенну рідера. Далі відбувається зчитування даних рідером та їх ретрансляція прикладній програмі, що виконується на комп'ютері через певні канали зв'язку (мережеве або послідовне з'єднання). Після цього програма використовує отримані відомості для ідентифікації об'єкта, піднесеного до рідера. Крім цього, програма може оновлювати інформацію в базі даних про місце розташування об'єкта, надсилати сигнал тривоги персоналу або просто ігнорувати дані. RFID-система – це набір компонентів, що становить єдине ціле, який реалізує будь-яке RFID-рішення. Вона включає такі компоненти: мітка, рідер, антенна рідера, система хост-комп'ютера і програмного забезпечення. RFID-система підтримує двосторонній обмін інформацією від рідерів до автоматизованої інформаційної системи та від автоматизованої інформаційної системи до рідерів [6].

Далі наведено порівняльний аналіз технологій штрихового кодування та RFID (табл. 1).

Таблиця 1 – Порівняльний аналіз основних техніко-економічних показників технологій штрихового кодування та RFID

Параметр	RFID	Штрихове кодування
Підтримка даних	Багаторазовий перезапис	Є статичним і не підлягає зміні
Видимість для зчитування даних	Не потребує	Завжди необхідна
Дистанція зчитування	Від 9 м до 91 м	До 9 м
Об'єм зберігання даних	Майже необмежений	До 3750 символів даних
Зчитування міток	Декілька міток за дуже короткий проміжок часу з використанням антиколізійної функції	Один штрих-код
Стійкість до впливу навколишнього середовища	Висока	Легко пошкоджується
Інтелектуальне	Застосовується для	Є лише засобом

застосування	виконання інших задач окрім зберігання та перенесення даних (спостереження за характеристиками навколишнього середовища)	зберігання даних
Точність зчитування	100 %	Від 90 % до 98 %
Вплив типу матеріалу	Важко зчитується на метали та деяких рідинах в ультрависокочастотному та мікрохвильовому діапазонах	Можна розміщувати на об'єкт майже із будь-якого матеріалу
Міжнародні правові обмеження	Не стандартизовано загальні робочі діапазони частот для роботи RFID у світовому масштабі. Частотні діапазони в одній країні можуть бути юридично несумісними в іншій країні	Для частот світлового діапазону не існує міжнародних обмежень
Ступінь зрілості технології	Технологія, що зароджується, невелика кількість типів прикладних систем, відсутність глобальних стандартів	Існує більше 30 років, широко використовується більше 50 стандартів, більшість з яких – по всьому світу
Вартість	Висока	Низька

IV. РЕЗУЛЬТАТИ

Порівнюючи технологію штрихового кодування і технологію RFID в логістичних процесах, було проведено аналіз порівняльних характеристик основних техніко-економічних показників. За результатами цього аналізу можна зробити висновки щодо порівняння вищезгаданих технологій за наступними параметрами:

– швидкість ідентифікації: використання RFID дозволяє ідентифікувати товари практично миттєво, в той час як штрихове кодування вимагає фізичного контакту і часу на сканування. Це робить RFID більш ефективним для ситуацій, де швидкість грає ключову роль, наприклад, у великих логістичних центрах;

– точність ідентифікації: RFID має вищу точність ідентифікації, оскільки воно не вимагає прямого видимого контакту і може працювати в умовах, де штрихові коди можуть бути пошкоджені або непрочитані;

– витрати на впровадження: штрихове кодування має менші витрати на впровадження, оскільки воно вже широко використовується та не вимагає спеціалізованого обладнання;

– витрати на обслуговування: витрати на обслуговування RFID включають в себе підтримку системи та заміну RFID-міток, що може бути вищими в порівнянні з обслуговуванням штрихових кодів;

– масштабованість: RFID може бути більш ефективним в великих обсягах, де потрібно відстежувати багато товарів одночасно в реальному часі. Штрихові коди можуть бути менш ефективними в таких умовах.

Також за результатами дослідження запропоновано структурну схему складської логістики з використанням RFID (рис. 1).

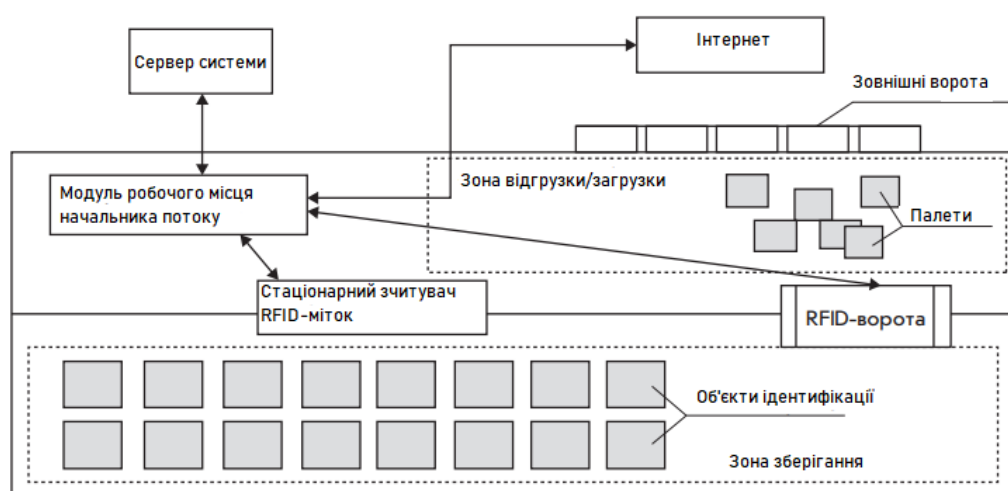


Рисунок 1 – Структурна схема складської логістики з використанням RFID

Для реалізації технології RFID у логістичному складі була створена структурна схема, яка включає наступні компоненти:

– модуль робочого місця начальника потоку: цей модуль включає в себе спеціалізовану робочу станцію, де начальник потоку може контролювати та відстежувати потік товарів;

- стаціонарний зчитувач RFID-міток: стаціонарний зчитувач встановлений на ключових точках складу і призначений для ідентифікації товарів, які проходять через нього;
- RFID-ворота: вони розташовані на вході та виході зі складу і дозволяють автоматично ідентифікувати товари, які в'їжджають та виїжджають;
- сервер системи: центральний сервер системи забезпечує збір та обробку даних, а також інтеграцію з іншими системами;
- зона завантаження/розвантаження: ця зона призначена для навантаження та розвантаження товарів на палети;
- зона зберігання: в даній зоні розміщені товари, які підлягають ідентифікації та відстеженню з використанням RFID;
- мережа Інтернет: мережа Інтернет забезпечує зв'язок модуля робочого місця начальника потоку з віддаленими системами.

Структурна схема з використанням RFID дозволяє забезпечити ефективну ідентифікацію та відстеження товарів на складі в режимі реального часу, що сприяє оптимізації логістичних процесів та зниженню помилок в управлінні складською логістикою.

V. ОБГОВОРЕННЯ

Отримані результати дослідження порівняння ефективності технології штрихового кодування і технології RFID у логістичних процесах відкривають цінну інформацію щодо можливостей і обмежень кожної з цих технологій та їх впливу на логістичні процеси. В цьому розділі пропонується інтерпретація результатів, розглядаються ключові аспекти та нюанси дослідження, а також обговорюються їх можливі наслідки та обмеження, а саме:

1. Швидкість та точність ідентифікації. Результати підтверджують, що RFID переважає штрихове кодування в плані швидкості та точності ідентифікації товарів. Це особливо важливо в сучасних умовах, де логістичні виробничі процеси вимагають великої точності та ефективності. З цього випливає, що RFID може забезпечити більш високу продуктивність та знизити ризик помилок у процесах ідентифікації товарів.

2. Витрати на впровадження та обслуговування. Штрихове кодування виграє в аспекті витрат на впровадження через свою доступність та не потребує значних інвестицій. Проте, результати також показали, що витрати на обслуговування RFID можуть бути вищими через необхідність підтримки системи та заміни RFID-міток. Ця інформація важлива для підприємств, які розглядають вибір технології.

3. Масштабованість. Один з важливих висновків полягає в тому, що RFID може бути більш ефективним в великих обсягах товарів і ситуаціях, де потрібно відстежувати їх у режимі реального часу. Це важливо для логістичних компаній та виробників, які оптимізували свої процеси для великих обсягів виробництва та постачання.

4. Специфіка робочого середовища. Також було виявлено, що обидві технології мають свої обмеження в специфічних робочих умовах. Штрихове кодування може бути менш ефективним в умовах, де штрихові коди піддаються пошкодженню. RFID, з іншого боку, може мати обмеження в металевих або вологих середовищах.

5. Перспективи комбінування технологій. Результати дослідження вказують на потенціал комбінування обох технологій в одній системі, щоб використовувати переваги кожної з них. Це може допомогти зменшити ризики та збільшити ефективність в ідентифікації та відстеженні товарів.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Необхідно враховувати, що це дослідження проводилося в певних умовах та за обмеженими обставинами. Результати можуть варіюватися в залежності від конкретних умов виробництва та логістики. Крім того, розвиток технологій може призвести до змін у порівняльних показниках обох технологій у майбутньому. Можна розглянути можливість включення додаткових технологій і методів ідентифікації товарів, таких як QR-коди, Bluetooth-мітки, або інші RFID-технології. Це допоможе дати більш об'єктивну оцінку доступних варіантів. Також для отримання більш достовірних результатів можна розглянути можливість проведення дослідження в реальних виробничих або логістичних умовах з використанням промислового обладнання та систем ідентифікації.

VII. ВИСНОВКИ

У даному дослідженні було проведено порівняльний аналіз ефективності технології штрихового кодування та технології RFID в логістичних виробничих процесах. Головною метою було визначити, яка з цих технологій є більш ефективною та придатною для застосування в різних галузях виробництва та логістики. Гіпотеза дослідження полягала в тому, що RFID може бути більш ефективною технологією ідентифікації та відстеження товарів у порівнянні з штриховим кодуванням. Отримані результати підтвердили цю гіпотезу.

Результати дослідження показали, що RFID забезпечує більшу швидкість та точність ідентифікації товарів, що робить його більш придатним для вимог сучасних логістичних виробничих процесів. Штрихове кодування може бути менш витратним на впровадження, але обслуговування RFID може вимагати вищих витрат у довгостроковій перспективі. Великі логістичні центри та виробничі підприємства можуть здобути більшу користь від застосування RFID завдяки його здатності ефективно відстежувати великі обсяги товарів в реальному часі. Важливо враховувати умови робочого середовища при виборі технології. RFID може бути більш придатним для умов, де штрихові коди піддаються пошкодженню.

Отже, дослідження підтвердило гіпотезу, що RFID може бути більш ефективною технологією ідентифікації та відстеження товарів у порівнянні з штриховим кодуванням. Результати дослідження можуть бути корисними для підприємств, які розглядають вибір технології для оптимізації своїх логістичних виробничих процесів. Однак важливо враховувати, що вибір технології повинен базуватися на конкретних вимогах та умовах виробництва. Дослідження також вказує на можливості подальшого розвитку інноваційних рішень та комбінування різних технологій для покращення логістичних процесів і оптимізації виробництва.

ЛІТЕРАТУРА

- [1] J. Zhou and J. Shi, "RFID localization algorithms and applications—a review", *J. Intell. Manuf.*, vol. 20, no. 6, pp. 695–707, Aug. 2008
- [2] X. Zhang, Q. Dong, and F. Hu, "Applications of RFID in logistics and supply chains: An overview", in *Inernat. Conf. Logistics Eng. Manage. 2012*, Chengdu, China. Reston, VA: Amer. Soc. Civil Engineers, 2012
- [3] L. Ruiz-Garcia and L. Lunadei, "Monitoring cold chain logistics by means of RFID", in *Sustainable Radio Frequency Identification Solutions*. InTech, 2010
- [4] D. Hanny, M. A. Pachano, L. G. Thompson, and J. Banks, *RFID Applied*. Wiley, 2007

- [5] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID", *Wireless Netw.*, vol. 10, no. 6, pp. 701–710, Nov. 2004
- [6] A. Astigarraga *et al.*, "A 21 m operation range RFID tag for "pick to light" applications with a photovoltaic harvester", *Micromachines*, vol. 11, no. 11, p. 1013, Nov. 2020

Мобільний застосунок для вимірювання земельних ділянок

Мерлак Ігор

Кафедра комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О. М. Бекетова
Харків, Україна

Анотація—В представленому матеріалі розглядається розробка мобільного застосунку для вимірювання земельних ділянок, за допомогою якого користувач може отримати результати розрахунків периметру або площі земельної ділянки; відстані між заданими точками; побачити схему території та, за необхідністю, вивести цю схему на карту.

Ключові слова—Мобільний застосунок, Об'єктно-орієнтоване проектування, Земельна ділянка, Циклічний список.

I. ВСТУП

Швидкий розвиток функціоналу мобільних пристроїв призвів до розробки прикладних застосунків, які можуть бути застосовані у різних сферах практичної діяльності людини. Розширення культури і використання смартфонів створило нову гілку можливостей для користувачів. Особливо тих, хто не міг дозволити собі дороге професійне обладнання для таких завдань, як вимірювання великих площ. До епохи автоматизації вимірювання площі та периметру земель виконувалися дуже примітивними інструментами, такими як вимірювальна стрічка або рулетка. Це призводило до величезних витрат часу та великих похибок у розрахунках.

В даний час робити вимірювання можна робити швидше, точніше і набагато простіше. GPS-системи [1] дають можливість вимірювати площу максимально точно, єдиною проблемою є неймовірно висока вартість обладнання, яка може сягати десятків тисяч євро. Зрозуміло, що такі витрати були недосяжними для малого та середнього бізнесу. Великі зміни відбулися коли на ринок вийшли доступні смартфони з підтримкою GPS. Розробники мобільних застосунків пропонують різноманітні інструменти для підвищення продуктивності, в тому числі й для вимірювання. Завдання, які раніше могли бути виконані лише спеціалізованими підприємствами або невеликою кількістю фахівців, тепер легко виконуються за допомогою смартфонів.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Використання мобільних пристроїв в житті сучасної людини не викликає жодних сумнівів. Сьогодні у світі налічується близько 3,5 мільярдів користувачів смартфонів. Google Android і Apple iOS разом становлять 99% світового ринку мобільного зв'язку. Відповідно, мобільні застосунки стали невід'ємною частиною нашого повсякденного життя. І про це свідчить статистика. За 2020 рік було близько 218 мільярдів завантажень мобільних застосунків. Кількість завантажень у 2022 році складала більше 200 млрд. А згідно з останніми дослідженнями data.ai перша половина 2023 року свідчить про повернення високих темпів зростання: споживчі витрати становили рекордні \$67,5 млрд, а кількість завантажень застосунків у всьому світі в iOS та Google Play сягнула 76,8 млрд. У першому півріччі загальні споживчі витрати зросли на 5,3% порівняно з аналогічним періодом минулого року, зокрема на застосунки на 16%. Кількість завантажень також зросла на 3,2% [2]. Частка мобільного трафіку у 2022 році становить більше 60% проти 10% у 2011. Дохід на ринку

мобільних застосунків сягне 420 млрд. дол у 2023 році, і до 2026 очікується його збільшення до 542 млрд. [3]. Ці статистичні дані говорять про непохиле зростання використання та зацікавленості в збільшенні мобільних застосунків за різними напрямками.

III. МЕТОДИ

Для визначення основних функцій мобільного застосунку була побудована функціональна модель системи, що представлена за допомогою UML діаграми використання [4] та виділені її основні актори. Як і в більшості мобільних застосунків, актором системи є її користувач. Основними функціями системи є обробка проєктів, що містять інформацію про земельні ділянки. Найважливішою є функція створення проєкту, яка передбачає вибір типу проєкту та додавання точок, що характеризують земельну ділянку. Функція «Обрахунок проєкту» реалізує обчислення геометричних характеристик земельної ділянки або ламаної, для чого виконується перетворення координат. Функція «Перегляд проєкту» дозволяє користувачу переглянути координати та зображення точок з створеного або завантаженого проєкту, виконати графічне відображення проєкту на мапі або схемі, а також побачити результати розрахунків. Функції роботи з базою даних дозволяють користувачу виконувати збереження нових проєктів та завантаження раніше створених проєктів з зовнішнього сховища.

Процес вибору алгоритмів для розв'язання задачі розглянуто за допомогою UML діаграми діяльності користувача мобільного застосунку. Першим кроком роботи з системою є обрання типу проєкту: земельна ділянка або ламана лінія. Далі користувачем виконується циклічне додавання у проєкт необхідної кількості кутових точок. Додавання кожної точки включає отримання її GPS-координат та фотозйомку місця розташування точки. Фотозйомка місця розташування точки дозволить користувачу у майбутньому зорієнтуватися щодо положення земельної ділянки на місцевості. На наступному кроці виконуються дії щодо перетворення GPS-координат точок у декартову систему, що дозволяє в подальшому виконати обчислення геометричних характеристик ділянки та відображення її схеми у декартовій системі координат. Після виконання розрахунків користувач може відобразити проєкт у вигляді схеми або на мапі, а також переглянути характеристики та зображення точок проєкту.

Для розробки мобільного застосунку для вимірювання земельних ділянок було проведено аналіз предметної області. На основі проведеного аналізу було визначено вхідні дані для роботи мобільного застосунку. Вхідними даними, якими задається земельна ділянка, є інформація про точки, а саме значення їх географічних координат – довготи та широти у форматі GPS (тип даних – дійсне число). До кожної точки можна прив'язати її фотографічне зображення, яке потім може бути виведено у мобільному застосунку. Один проєкт може містити до 100 точок. Обмеження на розмір та формат географічних координат встановлюються відповідно до стандарту всесвітньої системи координат WGS 84. Під час збереження проєкту у базу даних користувачем задається ім'я проєкту у вигляді текстового рядка, довжиною до 255 символів.

Вихідними даними застосунку є:

- розрахункові значення периметру та площі земельної ділянки або довжини ламаної лінії (дійсні числа);
- задані географічні GPS-координати точок переводяться у декартову систему координат XY (тип даних – дійсне число) і можуть вважатися проміжними даними;
- збереження усіх вхідних параметрів проєкту в базі даних;
- побудова схеми території та відображення її на Google-мапі.

Для організації збереження і подальшого завантаження вхідних даних проєкту з зовнішнього сховища було розроблено базу даних, структура якої наведена на ER-діаграмі (рис. 1).

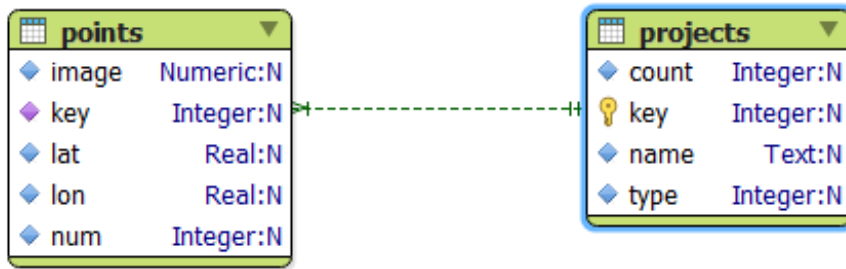


Рисунок 1 – ER-діаграма бази даних

Розроблена база даних складається з двох таблиць, які взаємодіють одна з одною у відношенні «один до багатьох».

На основі проведеного аналізу предметної області було виконано об'єктно-орієнтоване проєктування системи та побудована її об'єктно-орієнтована модель [5]. Під час проєктування було виділено 4 класи та три користувацьких типи даних, які наведено на діаграмі (рис. 2).

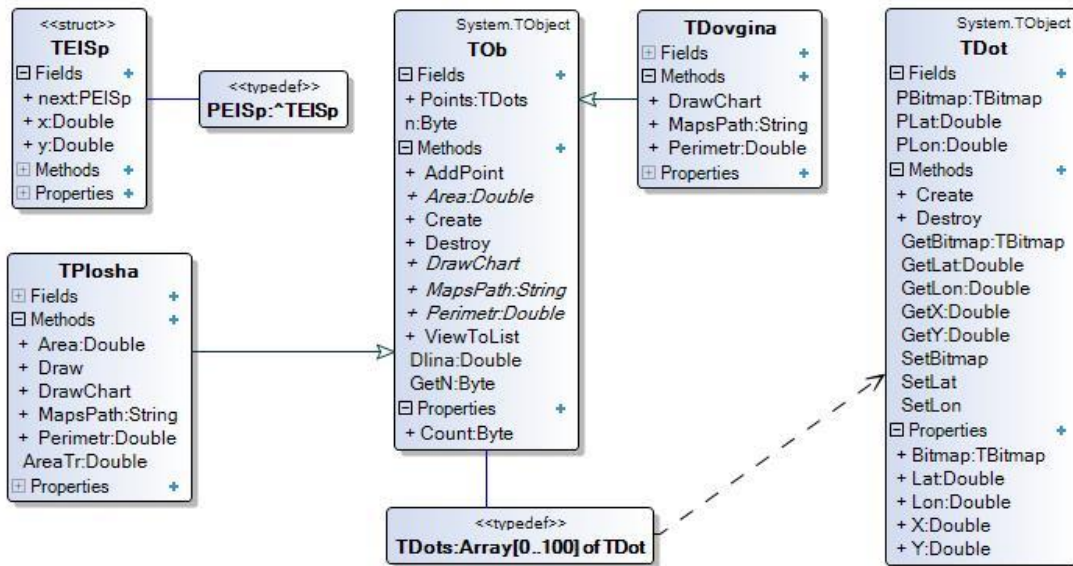


Рисунок 2 – Діаграма класів

Для обчислення площі багатокутника, який уявляє собою земельну ділянку, використовується метод трапецій. З метою послідовного обходу вершин багатокутника та формування трапецій багатокутник уявляється у вигляді структури даних однопов'язаний циклічний список, як такої, що відповідає графічному відображенню багатокутника. Для роботи з циклічним списком були розроблені 3 алгоритми: алгоритм додавання нового елемента у циклічний список, алгоритм перегляду циклічного списку та алгоритм видалення лінійного списку.

Для реалізації методів відображення точок проєкту на Google-мапах був розроблений алгоритм генерації web-запитів.

IV. РЕЗУЛЬТАТИ

Як результат представленої роботи можна визначити наступне. Проведено аналіз предметного середовища, в результаті якого були визначені фактори та необхідні умови функціонування мобільного застосунку для вимірювання земельних ділянок, за допомогою якого користувач може отримати результати розрахунків периметру або площі земельної ділянки; відстані між заданими точками; побачити схему території та, за необхідністю, вивести цю схему на карту. Проаналізована предметна область, на основі проведеного

аналізу визначені вхідні та вихідні дані системи, що проектується, проведено об'єктно-орієнтоване проектування системи та побудована її об'єктно-орієнтована модель. На основі даних, отриманих в процесі аналізу, було сформульовано відповідні математичні методи обчислення характеристик геометричних фігур. З метою послідовного обходу вершин багатокутника та формування трапецій багатокутник уявляється у вигляді односпрямованого циклічного списку. Для реалізації методів класів системи було побудовано наступні алгоритми: додавання елемента у циклічний список; перегляду циклічного списку; видалення лінійного списку; генерації web-запитів.

V. ОБГОВОРЕННЯ

Розроблена об'єктно-орієнтована модель мобільного застосунку для вимірювання земельних ділянок може бути використана в поєднанні з геодезичною зйомкою для фіксування даних. Для цього достатньо додати поле в мобільний застосунок для внесення результатів геодезичної зйомки для подальших розрахунків кількості землі, яку треба, припустимо, зняти для вирівнювання ділянки. Також є можливість використання такого мобільного застосунку в сільсько-господарському секторі для побудови схеми, припустимо, ділянок, засіяних різними культурами.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Наступним етапом в розробці мобільного застосунку є програмна реалізація зі створенням програмного засобу, відповідно до побудованої об'єктно-орієнтованої моделі, з детальним керівництвом користувача.

VII. ВИСНОВКИ

Як загальний висновок можна визначити наступне. В роботі обґрунтовано актуальність розробки – використання мобільного застосунку як важливого та необхідного засобу для покращення продуктивності. Під час реалізації поставленої задачі для уявлення багатокутника використовується структура даних пов'язаний односпрямований циклічний список. Перевагами використання однопов'язаного циклічного списку перед класичними лінійними послідовними структурами даних є: відсутність вираженого початку та кінця списку; можливість використання будь-якого вузла у якості початкового; легкість видалення будь-якого вузла списку, яка не потребує зсуву усіх елементів; легкість додавання будь-якого вузла списку, яка не потребує зсуву усіх елементів.

ЛІТЕРАТУРА

- [1] Richharia, Madhavendra; Westbrook, Leslie David. *Satellite Systems for Personal Applications: Concepts and Technology*. John Wiley & Sons, 2011, p. 443
- [2] Мобільний ринок 2023 року. Дата зварнення: 05 лист. 2023. [Онлайн]. Доступно: <https://speka.media/shho-z-mobilnim-rinkom-u-2023-roci-zvit-dataai-p0j4I9>
- [3] EasternPeak. Дата зварнення: 05 лист. 2023. [Онлайн]. Доступно: <https://careers.easternpeak.com/blog/mobile-app-development-trends/>
- [4] Stanislaw Wrycza, Bartosz Marcinkowski. *Towards a Light Version of UML2.X. Appraisal and Model*, 2007, p. 192
- [5] Grady Booch, Robert A. Maksimchuk, Michael W. Engle, Bobbi J. Young, Ph.D. Jim Conallen, Kelli A. Houston. *Object-oriented analysis and design with applications*. 3rd ed. Addison-Wesley, 2002, p. 560

Інтелектуальне управління міською інформаційною інфраструктурою

Наталія Братерська

*Кафедра комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О. М. Бекетова
Харків, Україна*

Анотація — Робота присвячена аналізу розвитку інтелектуальних технологій управління та впровадження їх у міське управління, огляду ключових компонентів *smart-міста*, методів та інструментів оптимізації функціонування та покращення міської інфраструктури із залученням інтелектуального управління.

Ключові слова — *Інтелектуальне управління, Міська інфраструктура, Інформаційна інфраструктура, Smart-місто, Інтернет речей, Штучний інтелект.*

I. ВСТУП

Міська інформаційна інфраструктура є невід'ємною частиною сучасного життя, і в останні десятиліття вона стала ще більш важливою для функціонування міст у цифрову епоху. Швидкий технологічний прогрес і зростаюча кількість жителів у містах створюють великий попит на інноваційні підходи до управління цією структурою. Інформаційна інфраструктура – це комплекс інформаційних технологій, систем та інфраструктурних ресурсів, які використовуються для забезпечення різних аспектів функціонування та управління містами та муніципалітетами. Управління міською інфраструктурою взагалі може використовувати інформаційні технології, але не обов'язково засноване на них. В свою чергу, інтелектуальне управління інфраструктурою міст передбачає використання сучасних технологій, таких як Інтернет речей (IoT), штучний інтелект (AI) та машинне навчання (ML), для автоматизації та оптимізації управління міськими системами. Smart-місто – це місто, яке використовує технології та методи інтелектуального управління для досягнення конкретних цілей. Інтелектуальне управління може допомогти містам підвищити ефективність використання ресурсів, зменшити витрати та покращити якість життя, що відображається загальним станом та задоволенням людей індивідуальними та соціальними життєвими умовами.

II. АНАЛІЗ

На сьогоднішній день, перед містами у всьому світі постають серйозні виклики, пов'язані з зростанням населення та погіршенням стану навколишнього середовища. У відповідь на ці виклики, багато міст звертаються до розвитку "smart-міст", які базуються на використанні штучного інтелекту для створення більш стійкого та життєздатного міського середовища. Smart-міста використовують AI для збору та аналізу даних з різних джерел, таких як дані про погоду, якість повітря, схеми руху та споживання енергії. Ці дані можуть бути використані для автоматизованого прийняття рішень і додатків "smart-міста", які підвищують стійкість, що визначається здатністю системи до збереження стабільності в умовах невизначеності, і якість життя. Інтелектуальне управління міською інформаційною інфраструктурою у свою чергу може принести ряд переваг містам, зокрема: підвищити ефективність роботи міських систем, таких як транспорт, енергія, водопостачання та управління відходами; зменшити витрати на утримання інфраструктури; покращити якість життя в містах, наприклад, зменшивши забруднення повітря і шум. Але водночас інтелектуальне управління може бути пов'язане з рядом викликів, такими як: дорога вартість впровадження; складні недосконалі технології; впровадження може викликати соціальні та політичні конфлікти інтересів різних груп населення [1-3].

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

Згідно з дослідженням консалтингового агентства Navigant Research, smart-міста включають такі ключові компоненти[4]:

- Розумна енергія: комплекс рішень у сфері енергопостачання та енергозбереження (програми управління попитом, енергоефективність, інтеграція відновлюваних джерел енергії);

- Розумна вода: управління водними ресурсами (модернізація систем водопостачання, моніторинг галузевого водоспоживання, екологічна безпека, системи запобігання паводкам);

- Розумні будівлі: включає окремі будівельні конструкції та установки, де всі інженерні та інформаційні системи об'єднані та інтегровані в єдину систему управління (Building Management System: BMS). Такі системи можуть, наприклад, обігрівати будівлю в необхідні робочі години відповідно до кількості людей в будівлі, регулювати потужність і якість повітря в системах вентиляції або автоматично переходити в режим енергозбереження, коли в будівлі немає людей;

- Розумний транспорт: інтелектуальні транспортні та логістичні системи можуть бути побудовані для моніторингу та управління дорожнім рухом, що дозволяє контролювати ціноутворення на дорогах, реагування на надзвичайні ситуації та сигналізацію. Інтелектуальні системи паркування та сповіщення про зупинки громадського транспорту також часто розглядаються в цій сфері;

- Розумне урядування: використання інформаційних технологій для надання державних послуг широкому колу осіб та оптимізації роботи різних секторів.

III. МЕТОДИ

Міська інфраструктура стає все більш складною і вимагає інноваційних підходів для її керування. Інтелектуальне управління – це один з таких підходів. Серед методів та інструментів оптимізації функціонування та покращення інфраструктури міста, можна виділити наступні [1,5]:

1. Використання сенсорів та Інтернет речей (Internet of Things):

Один із ключових методів інтелектуального управління – це встановлення сенсорів і систем Інтернет речей (IoT) для моніторингу та збору даних з різних аспектів міської інфраструктури. Сенсори можуть вимірювати рух транспорту, рівень води, якість повітря, витрати енергії тощо. Ці дані надходять в реальному часі і стають важливим джерелом для прийняття рішень.

2. Великі дані та аналітика:

Обробка та аналіз великих обсягів даних (Big Data) використовуються для витягнення цінної інформації з сенсорних та інших даних. Методи аналітики даних, включаючи машинне навчання та штучний інтелект, дозволяють виявляти патерни та робити прогнози щодо стану інфраструктури та ресурсів.

3. Системи геоінформацій (GIS):

Геоінформаційні системи використовують географічну інформацію для аналізу та візуалізації міської інфраструктури. Вони допомагають у розміщенні і плануванні інфраструктурних об'єктів, моніторингу їхнього стану та координації дій в реальному часі.

4. Управління транспортом:

Інтелектуальне управління транспортом включає в себе системи контролю руху, оптимізацію маршрутів, управління світлофорами та зменшення транспортних заторів. Це сприяє покращенню мобільності та зменшенню забруднення довкілля.

5. Енергозбереження та сталість:

Використання інтелектуальних систем управління дозволяє ефективно використовувати енергію, впроваджувати відновлювані джерела енергії та зменшувати викиди CO₂ для забезпечення сталого розвитку.

6. Ефективне водопостачання та водовідведення:

Методи інтелектуального управління проводити моніторинг якості води, виявлення витоків та аварій в системах водопостачання та водовідведення.

7. Залучення громадян:

Міські платформи та мобільні застосунки дозволяють громадянам бути активними учасниками процесу збору даних, сповіщення про проблеми та взаємодії з міською владою.

8. Кіберзахист:

Захист від кіберзагроз стає все важливішим в контексті підключення міської інфраструктури до Інтернету та обміну даними.

IV. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У подальшому напрям досліджень планується зосередити саме та методах та інструментах інтелектуального управління якістю міської інфраструктури загалом та окремо міської інформаційної інфраструктури. Результати таких досліджень забезпечать здійснити аналіз існуючих методів та на їх основі розробити нові підходи, які посприяють ефективно покращити та вдосконалити якість інформаційної інфраструктури міста.

V. ВИСНОВКИ

Розвиток інтелектуальних технологій та впровадження їх у міське управління є ключовими кроками у напрямку створення більш стійких та якісних міст для населення. Проте, важливо враховувати всі виклики та питання, пов'язані з цим процесом, і долати їх шляхом спільних зусиль влади, громадян та бізнесу.

ЛІТЕРАТУРА

- [1] “К. Маркевич, Smart-інфраструктура у сталому розвитку міст: Світовий досвід та перспективи України. Київ: Вид-во “Заповіт”, 2021. Р. 400 [Online]. Available: <https://razumkov.org.ua/uploads/other/2021-SMART-CYTI-SITE.pdf>
- [2] “Easy park.” This is EasyPark - EasyPark Group. Accessed: Nov. 5, 2023. [Online]. Available: <https://www.easyparkgroup.com/studies/cities-of-the-future/en/У>
- [3] “World cities report 2022.” UN-Habitat - A Better Urban Future | UN-Habitat. Accessed: Nov. 5, 2023. [Online]. Available: <https://unhabitat.org/wcr/>
- [4] Main Site - Smart city: технології «розумного міста» та їх цільове призначення.” Головна | Е-Україна Е-Ukraine. Accessed: Nov. 5, 2023. [Online]. Available: <https://eukraine.org.ua/ua/news/smart-city-tehnologiyi-rozumnogo-mista-ta-yih-cilove-priznachennya>
- [5] Gramchuk, Marina, Nikitenko, Vitalina. Present trends and prospects of smart city development. Humanities studies: Collection of Scientific Papers / Ed.V. Voronkova. Zaporizhzhia : Publishinghouse “Helvetica”, 2023. 14 (91). P. 35–41. doi: <https://doi.org/10.32782/hst-2023-14-91-04>

Методи автоматизації та оптимізації побудови навчального матеріалу в інтелектуальних адаптивних вебсистемах самоосвіти

Марія Дутчак, Андрій Аннич, Олег Козич

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація— У статті проведено аналіз сучасного стану та перспектив розвитку штучного інтелекту в вебсистемах самоосвіти та описано запропонований метод автоматизованої побудови, оптимізації та вибору кращої за певним критерієм адаптивної траєкторії навчання шляхом автоматизації та оптимізації побудови навчального матеріалу в інтелектуальних адаптивних вебсистемах самоосвіти.

Ключові слова— *інтелектуальні адаптивні вебсистеми самоосвіти, штучний інтелект, адаптивна траєкторія навчання, оптимізація навчального матеріалу.*

I. ВСТУП

Розвиток мережі Інтернет, хмарні обчислення, технології великих даних і значні досягнення в галузі штучного інтелекту (ШІ) змінили освіту. Останніми роками з'явилося багато публікацій, які описують більш досконалі навчальні системи з підтримкою штучного інтелекту, які набувають популярності завдяки своїй здатності надавати навчальний контент і адаптуватися до індивідуальних потреб здобувачів освіти. Проте, незважаючи на те, що ці сучасні системи навчання є корисними освітніми платформами, які дозволяють покращити якість надання освітніх послуг, існує багато проблем і потреб, вирішення яких змогло б підняти даного класу системи на якісно новий рівень надання освітніх послуг, зокрема у напрямку розвитку методів автоматизації побудови адаптивної траєкторії навчання із зменшенням потреби втручання у навчальний процес викладачів чи інженерів зі знань.

Функціонування інтелектуальних адаптивних вебсистем самоосвіти (ІАВСС) можна описати як постійне прийняття рішень щодо побудови адаптивного навчального процесу на основі накопичених знань та аналізу поточних ситуацій для досягнення поставленої мети. Одним із важливих критеріїв якості даних систем є можливість самонавчання та самовдосконалення в процесі проходження занять студентами [1].

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Групою авторів у складі Т. Kabudi, I. Pappas, D. Olsen [2] за результатами аналізу достатньо великої кількості наукових публікацій були виділені використовувані методи штучного інтелекту та аналізу даних. Найчастіше згадуваною у дослідженнях виявилась технологія байєсівських мереж. Наступною найбільш часто згадуваною технікою були нейронні мережі. Древа рішень, генетичні алгоритми та методи К-найближчих сусідів (KNN) також були ідентифіковані в цьому дослідженні. Далі йдуть дослідження, які ґрунтуються на векторних машинах (SVM) і байєсівському відстеженні знань (ВКТ), а також у дослідженні наведено багато інших менш часто згадуваних методик.

Проте, на думку багатьох авторів [3-5], які досліджували впровадження методів ШІ у реально діючі навчальні системи, більшість адаптивних систем навчання на практиці функціонують просто як звичайні навчальні платформи. Зазначається, що існують значні

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

розбіжності між тим, що можуть робити описані у літературі навчальні системи з підтримкою ШІ, і тим, як вони фактично впроваджені і апробовані на практиці.

Окрім того, існуючі освітні системи використовують, як правило, стандартизовані методи навчання, які не відповідають індивідуальним особливостям кожного здобувача освіти [6]. Це підкреслює необхідність розробки та використання нових методів штучного інтелекту під час побудови навчального процесу засобами ІАВСС, щоб дані системи навчання могли задовольняти різні потреби та особливості кожного із здобувачів освіти.

Отже, проведений аналіз стану розробки та впровадження методів та моделей побудови інтелектуального адаптивного навчання показав потребу в їх подальшому розвитку та впровадженню передових технологій ШІ. Зокрема актуальною є проблема автоматизації побудови індивідуалізованого навчання, що полягає в створенні і забезпеченні студентів навчальним матеріалом, який відповідає їх можливостям та меті навчання із врахуванням академічних вимог.

III. МЕТОДИ

У даній публікації пропонується метод побудови та вибору оптимальної адаптивної траєкторії навчання (АТН) в розширених базах навчальних матеріалів, в яких можлива наявність декількох навчальних блоків (НБ), що пояснюють одне і те ж поняття. Такі НБ можуть відрізнятися тими чи іншими параметрами, наприклад, рівнем зв'язності з іншими НБ та ключовими поняттями, складністю, доступністю викладу, актуальністю навчального матеріалу, тривалістю вивчення, вартістю і т.д.

Для вибору оптимальних параметрів АТН використовується оцінка ймовірностей досягнення поставленої мети при різних сценаріях навчання та оптимізація функції пристосованості засобами квантового генетичного алгоритму вищих порядків. Основою для формування керуючих впливів є можливості та мета навчання студента. [7]

Розглянемо метод побудови АТН. Завдання побудови АТН формулюється в такий спосіб: побудовано та заповнено необхідні слоти семантичної мережі навчального матеріалу (СМНМ) [8]. Задано множину ключових понять C_{key} . Потрібно побудувати оптимальний підграф СМНМ, що містить траєкторію переходу від вихідних понять (або від понять, які не мають входів, необхідних для засвоєння студентами) до кожного ключового поняття. Очевидно, що в загальному випадку задача має багато рішень. Серед них потрібно вибрати одне, оптимальне по заданому критерію.

Пояснення понять містяться в НБ. Для того, щоб було вивчене поняття C_i потрібно в траєкторії навчання мати хоча б один НБ, що пояснює це поняття. Іншими словами, щоб C_i -змінна мала значення true, потрібне прийняття значення true хоча б однієї з b -змінних тих НБ, в яких визначається C_i . З іншого боку, щоб НБ, який містить пояснення поняття C_i (нехай це НБ b_{ij}) був зрозумілий студенту, студент повинен знати або попередньо вивчити вхідні для НБ b_{ij} поняття. Змінні цих понять повинні мати значення true.

Отже, автоматизована побудова траєкторії навчання повинна здійснюватися з дотриманням наступного правила:

$$C_{key_i} = \prod_{j \in J_1} b_{ij} + \sum_{j \in J_2} r \in R_j b_{ij} \prod c_{jr} + \prod_{j \in J_3} b_{ij}(zk_{j,j+1}) = true, \quad (1)$$

де C_{key_i} – i -те ключове поняття; b_{ij} – j -й НБ, який є вхідним для i -го НБ, J_1 – множина НБ, які передують в ієрархії НБ блоку, в якому визначене C_{key_i} , для яких $zk_1 = 1$; J_2 – множина НБ, в яких визначене поняття C_{key_i} , R_j – множина вхідних для НБ b_{ij} понять; J_3 – множина НБ, які змістово-зв'язні з НБ, в якому визначене C_{key_i} , для яких $zk_3 > 0$.

Побудова навчальної траєкторії починається з вибору ключових понять. Цей вибір виконує викладач, орієнтуючись на академічні вимоги. При самоосвіті, студент самостійно вибирає ключові поняття, які є кінцевою метою навчання. Далі виконується циклічний обчислювальний процес вибору НБ на основі формули (1) для кожного ключового поняття з послідовною заміною в правих частинах формул c -змінних на b -змінні і нові c -змінні. Цей процес триває до тих пір, поки в крайніх правих частинах не буде змінних інших понять і НБ, окрім початкових, тобто тих, в яких відсутні вхідні поняття, які необхідно вивчити студенту.

Оптимізація і вибір однієї з побудованих траєкторій навчання виконується у відповідності до сформульованої мети навчання. Нехай метою навчання є засвоєння максимального обсягу навчального матеріалу курсу та досягнення максимального рівня знань Q при визначеній тривалості навчання T та переліку ключових компетенцій.

Функція пристосованості матиме вигляд:

$$Q = \sum_{i=1}^n C_i L_{zi_k}^t \rightarrow \max, \text{ при } \sum_{i=1}^n t_i \leq T, L_{zi_k}^t > d_i, \quad (2)$$

де $L_{zi_k}^t$ – найімовірніший рівень засвоєння навчального матеріалу (НМ) i -го заняття k -ї складності за час t деяким студентом, n – кількість занять, C_i – рівень важливості i -го заняття, d_i – мінімальне значення рівня засвоєння НМ i -го заняття, при якому заняття може бути зарахованим. $L_{zi_k}^t$ дорівнює рівню засвоєння НМ, для якого оцінка ймовірності, обчислена за формулою (2) буде найвищою.

Оцінка ймовірності того, що в момент часу t студент перебуває в стані L_j , визначається за формулою [9]:

$$p_j(t) = \sum_i p_i(0) p_{ij}^t, \quad (3)$$

де $p_i(0)$ – оцінка ймовірності перебування студента в i -му стані засвоєння; p_{ij}^t – оцінка умовних ймовірностей переходів із одного стану в інший для занять трьох рівнів складності.

Для оптимізації функції пристосованості було розроблено і використано квантовий генетичний алгоритм вищих порядків (QGA), який забезпечує глобальний пошук розв'язку при швидкій збіжності та невеликому розмірі популяції (випадковим чином згенерована множина можливих розв'язків, за допомогою якої досліджується область пошуку) [10].

IV. РЕЗУЛЬТАТИ

Розроблені моделі та методи були апробовані під час організації навчального процесу студентів спеціальностей 121 Інженерія програмного забезпечення та 113 Прикладна математика Прикарпатського національного університету імені Василя Стефаника (м. Івано-Франківськ) з дисциплін «Front-end розробка», «Програмування Інтернет», «Програмування мовою PHP» та «Веб-програмування».

Дослідження таких показників як «релевантність побудованої адаптивної траєкторії навчання» та «цілісність представлення адаптованого навчального матеріалу», проводилось шляхом опитування студентів з виставленням оцінки по кожному параметру за 100-бальною шкалою, і були високо ними оцінені.

Оцінку впливу на якість засвоєння НМ було проведено методом спліт-тестування (A/B тестування). Статистичні обчислення виконані на основі вибірки результатів навчання 3-ох груп зазначених вище спеціальностей у 2021/2022 н. р. та 3-ох інших груп за перший семестр 2022/2023 н. р. Порівняння вибірок студентів до експерименту проводилось з використанням t-критерію Стьюдента та критерію Пірсона χ^2 -квадрат (χ^2) [9].

V. ОБГОВОРЕННЯ

Аналіз даного дослідження показав, що під час організації навчання засобами розробленої системи, якість засвоєння НМ студентами збільшилась на 15%, тобто відбувся перехід студентів із категорій E і D в одну із категорій A, B, C, що свідчить про суттєвий вплив на рівень засвоєння студентів, які належать до даних категорій. Середнє арифметичне рівня засвоєння НМ у експериментальній групі збільшилось на 6,7% та на 5,2% у порівнянні відповідно із першою та другою контрольними групами.

Впровадження і використання розроблених методів побудови ІТН дозволяє підвищити якість автоматично організованого навчального процесу, раціональніше використовувати трудові і технічні ресурси закладів та систем освіти шляхом автоматизації та адаптації процесу оволодіння знаннями.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Перспективним напрямом подальших досліджень є розвиток і впровадження методів та засобів аналізу даних та інших методів штучного інтелекту з метою оптимізації і підвищення якості навчання студентів засобами ІАВСС.

VII. ВИСНОВКИ

Отже, представлене дослідження сприяє подальшому розвитку методів автоматизованої побудови навчального матеріалу із наявних навчальних блоків (модулів). Також було запропоновано удосконалену модель навчального матеріалу та удосконалений метод встановлення змістовних коефіцієнтів зв'язку між навчальними блоками, що дозволяє підвищити коректність та оптимальність автоматично сформованих варіантів навчальних траєкторій.

Запропоновано удосконалений метод автоматизованої побудови, оптимізації та вибору кращої за певним АТН, який дає можливість підвищити якість процесу оволодіння знаннями в інтелектуальних адаптивних вебсистемах самоосвіти. Даний метод дозволяє будувати АТН в залежності від параметрів моделі студента, виявленого рівня знань ключових та проміжних понять та академічних вимог, а також компактніше розміщувати більш зв'язні навчальні блоки, сприяючи там самим кращому їх засвоєнню.

1. J. Ong and S. Ramachandran, "An Intelligent Tutoring System Approach to Adaptive Instructional Systems," United States Army Research Institute for the Behavior and Social Sciences, CA, ADA437533, Sept. 2005. Date of access: November 1. 2023. [Online]. Available: <https://apps.dtic.mil/sti/pdfs/ADA437533.pdf>
2. T. Kabudi, I. Pappas and D. H. Olsen, "AI-enabled adaptive learning systems: A systematic mapping of the literature," *Comput. & Education: Artif. Intell. (CAEAI)*, vol. 2, 2021, ar. no. 100017. <https://doi.org/10.1016/j.caeai.2021.100017>
3. M. Hou and C. Fidopiastis, "A generic framework of intelligent adaptive learning systems: from learning effectiveness to training transfer," *Theor. Issues Ergonom. Sci.*, vol.18, no. 2, p. 167–183, 2017. <https://doi.org/10.1080/1463922X.2016.1166405>
4. G. Padron-Rivera, C. Joaquin-Salas, J.-L. Patoni-Nieves and J.-C. Bravo-Perez, "Patterns in Poor Learning Engagement in Students While They Are Solving Mathematics Exercises in an Affective Tutoring System Related to Frustration," *Lecture Notes Comput. Sci.*, vol. 10880, p. 169–177, 2018. https://doi.org/10.1007/978-3-319-92198-3_17
5. H. Xie, H.-C. Chu, G.-J. Hwang and C.-C. Wang, "Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017," *Comput. & Educ.*, vol. 140, p. 103599. <https://doi.org/10.1016/j.compedu.2019.103599>
6. M. Oliveira, A. Barreiras, G. Marcos, H. Ferreira, A. Azevedo та C. Vaz de Carvalho, "Collecting and Analysing Learners Data to Support the Adaptive Engine of OPERA, a Learning System for Mathematics," in *Special Session Analytics Educational Environ.*, Porto, Portugal, 21–23 Apr. 2017. SCITEPRESS - Sci. Technol. Publications, 2017. <https://doi.org/10.5220/0006389806310638>
7. M. Dutchak, M. Kozlenko, I. Lazarovych, N. Lazarovych, M. Pikuliak and I. Savka, "Methods and Software Tools for Automated Synthesis of Adaptive Learning Trajectory in Intelligent Online Learning Management Systems," in *Innovations in Smart Cities Applications*, Volume 4. Cham: Springer Int. Publishing, 2021, p. 206–217. https://doi.org/10.1007/978-3-030-66840-2_16
8. П. І. Федорук та М. С. Дутчак, "Побудова бази знань адаптивних систем дистанційного навчання на основі фреймової та продукційної моделей представлення знань," *Управляючі системи і машини (УСiМ)*, № 5, с. 3–10, 2012. <https://doi.org/10.15407/csc>
9. Я. Бродський, "Характеристики Статистика. Ймовірність. Комбінаторика," Київ: НК - Богдан, 2014.
10. V. M. Tkachuk, M. I. Kozlenko, M. V. Kuz, I. M. Lazarovych and M. C. Dutchak, "Optimization Based on Higher-Order Quantum Genetic Algorithm," *Electronic modeling*, vol. 41, no. 3, p. 43–58, Jun. 2019. <https://doi.org/10.15407/emodel.41.03.043>

Сертифікаційна модель автоматизованої програмної системи управління навчальним закладом WEBportal ПНУ

Руслан Запухляк, Микола Кузь, Микола Козленко, Микола Пікуляк,
Ігор Лазарович, Валерій Ткачук, Борис Незамай

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація—Доведено необхідність проведення сертифікації програмного забезпечення управління навчальним закладом. Побудована ієрархічна модель якості програмних продуктів, що встановлює взаємозв'язки між характеристиками, підхарактеристиками, властивостями якості, а також метриками вимірювання властивостей якості та ваговими коефіцієнтами. Розроблена модель якості програмного забезпечення АСУ WEBportal, що є власною розробкою Прикарпатського національного університету імені Василя Стефаника. Розроблена сертифікаційна модель якості даного програмного забезпечення, що встановлює чітку процедуру сертифікаційних досліджень програмного забезпечення АСУ WEBportal.

Ключові слова—*Модель якості, Ієрархічна модель якості, Сертифікаційна модель якості, Автоматизована система управління навчальним закладом.*

I. ВСТУП

Проведення освітньої діяльності в закладах вищої освіти здійснюється з використанням багатьох програмних засобів і систем. Для їх ефективного функціонування необхідно оцінювати їх відповідність нормативним документам, що встановлюють вимоги до таких програмних продуктів. Особливо це актуально для програмного забезпечення власної розробки навчальних закладів. Для забезпечення найкращої об'єктивної оцінки відповідності програмних засобів і систем вимогам нормативних документів є процедура їх сертифікації. Для ідентифікації показників якості програмних продуктів, формування вимог до них та критеріїв їх оцінки прийнято використовувати моделі якості.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Вимоги до якості систем і програмних засобів наведені в міжнародному стандарті [1]. Цей нормативний документ також містить моделі якості та сферу їх застосування. Стандарт [1] є нормативним документом, який замінює собою інший міжнародний стандарт [2] шляхом його оновлення. У стандарті [2] долучено ряд характеристик та підхарактеристик якості, у порівнянні із стандартом [1], а також об'єднанні два поняття: зовнішня та внутрішня якість в одне поняття: якість програмних продуктів.

Однак документ [2], а також [3] і [4] регламентують тільки загальну номенклатуру показників якості програмних продуктів, але не наводять перелік показників якості, які необхідно визначати при проведенні оцінки (сертифікації) того чи іншого програмного продукту.

Для вирішення цього питання в роботі [5] пропонується розробляти сертифікаційну модель якості програмних систем. В [5] також наведена концепція побудови такої моделі.

III. МЕТОДИ

В роботі [5] запропонована методика побудови загальної ієрархічної моделі якості, що записується у вигляді множини і, у нашому випадку, має наступний вигляд:

$$Q = \left\{ C_i \left\{ S_{ij} \left\{ P_{ijk} \left\{ M_{ijkl}, W_{ijkl} \right\}_{l=1}^{L_{ijk}} \right\}_{k=1}^{K_{ij}} \right\}_{j=1}^{J_i} \right\}_{i=1}^I, \quad (1)$$

де C_i – i -та характеристика; S_{ij} – j -та підхарактеристика i -тої характеристики; P_{ijk} – k -та властивість якості j -тої підхарактеристика i -тої характеристики; M_{ijkl} – метрики вимірювання властивостей якості; W_{ijkl} – вагові коефіцієнти, L_{ijk} , K_{ij} , J_i , I – кількість метрик вимірювання властивостей якості (починаючи з $l=1$), властивостей якості, підхарактеристик, характеристик відповідно.

IV. РЕЗУЛЬТАТИ

Запропонована концепція для побудови сертифікаційної моделі якості дає змогу розробити таку модель для певного програмного продукту з врахуванням вимог міжнародних стандартів [1, 3, 4]. На рис. 1 зображено модель якості програмного забезпечення АСУ WEBportal, що є власною розробкою Прикарпатського національного університету імені Василя Стефаника.

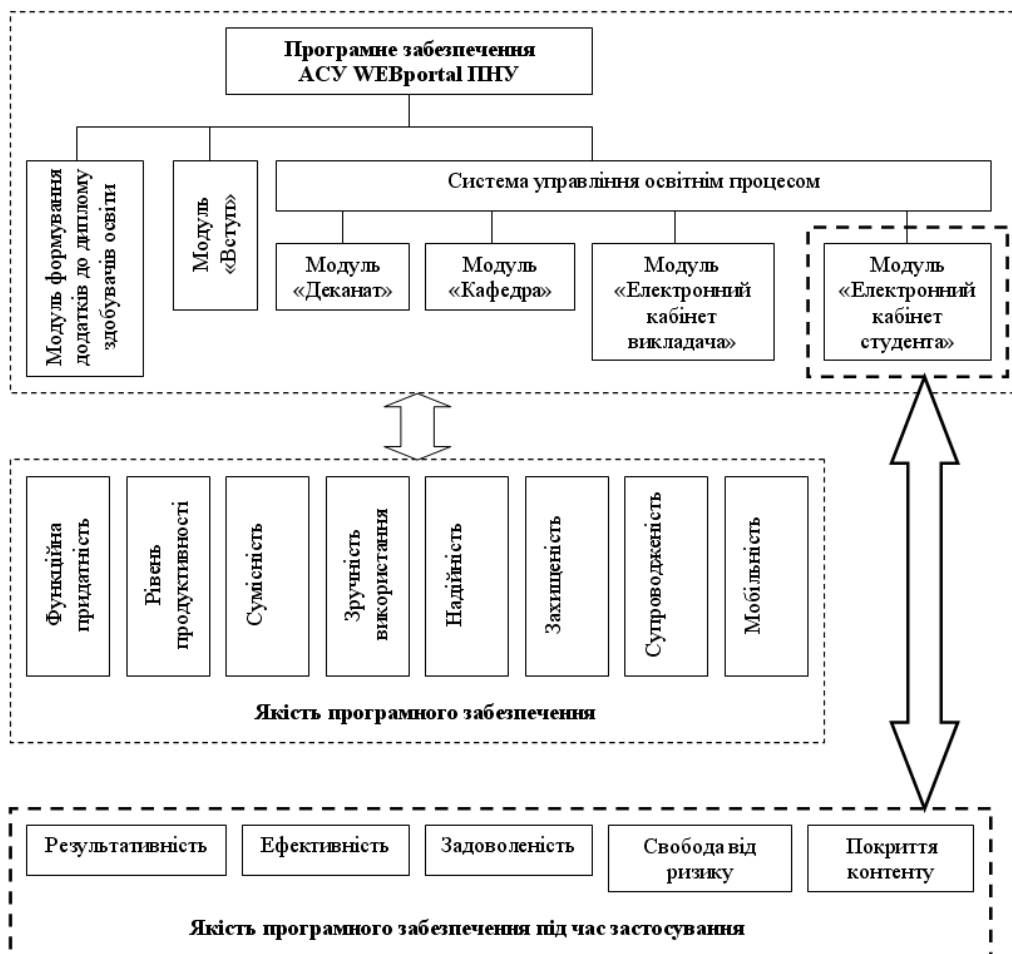


Рисунок 1 - Модель якості програмного забезпечення АСУ WEBportal

Модель якості (рис. 1) містить перелік модулів АСУ WEBportal, а також перелік характеристик якості програмного забезпечення, регламентованих міжнародним стандартом [3], та характеристик якості програмного забезпечення під час застосування, регламентованих в [4]. Модель (рис. 1) встановлює для кожного модуля перелік характеристик якості, які будуть оцінені під час проведення сертифікаційних випробувань. Як видно із рис. 1, унікальним є модуль «Електронний кабінет студента», який на відміну від інших модулів, є виключно користувацьким. Тому в цьому модулі застосовуються тільки характеристики якості програмного забезпечення під час застосування.

Ця модель є першим етапом для розроблення сертифікаційної моделі, структура якої наведена на рис. 2.

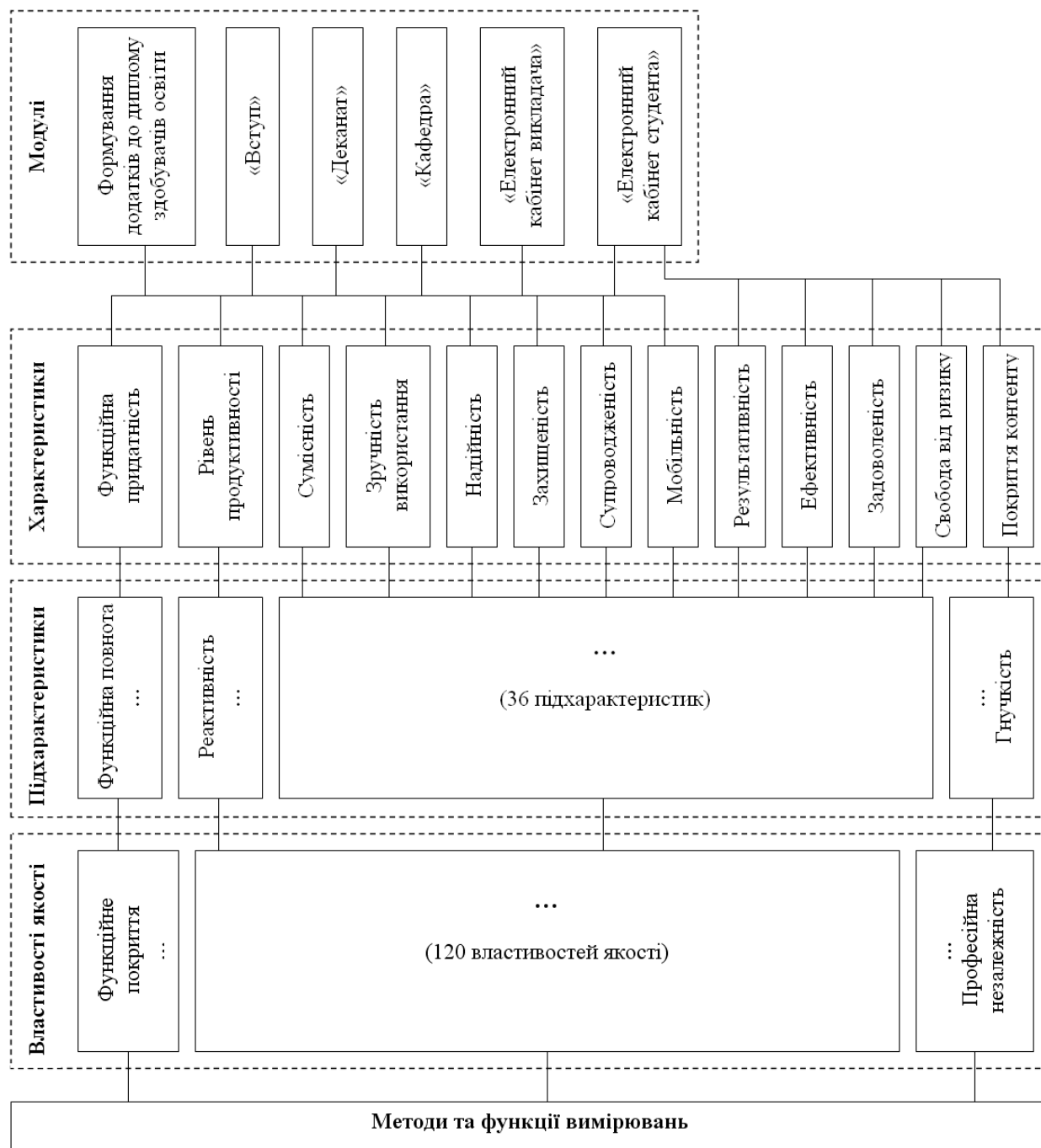


Рисунок 2 – Сертифікаційна модель якості програмного забезпечення АСУ WEBportal

V. ОБГОВОРЕННЯ

Розроблена сертифікаційна модель якості (рис. 2) встановлює чітку процедуру сертифікаційних досліджень програмного забезпечення АСУ WEBportal у розрізі визначення якісних показників кожного з модулів цього програмного продукту. Використання розробленої сертифікаційної моделі може бути основою для здійснення процедури сертифікації програмного забезпечення АСУ WEBportal чи іншого програмного забезпечення аналогічного призначення.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Предметом подальших наукових досліджень буде розроблення методології оцінки властивостей якості та вагових коефіцієнтів характеристик та підхарактеристик якості програмних продуктів.

VII. ВИСНОВКИ

Ефективність функціонування програмного забезпечення напряму залежить від його відповідності нормативним документам, що встановлюють вимоги до нього. Для забезпечення найкращої об'єктивної оцінки відповідності програмних засобів і систем під час їх сертифікації прийнято використовувати моделі якості. З цією метою побудована ієрархічна модель якості програмних продуктів, що встановлює взаємозв'язки між характеристиками, підхарактеристиками, властивостями якості, а також метриками вимірювання властивостей якості та ваговими коефіцієнтами; розроблена модель якості програмного забезпечення та сертифікаційна модель якості даного програмного забезпечення, що встановлює чітку процедуру сертифікаційних досліджень.

ЛІТЕРАТУРА

- [1] *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Systems and software Quality models*, ISO/IEC 25010,2011
- [2] *Software engineering – Product quality – Part 1: Quality model*, ISO/IEC 9126-1,2001
- [3] *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of quality in use*, ISO/IEC 25022,2016
- [4] *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of system and software product quality*, ISO/IEC 25023,2016
- [5] І. Райчев, О. Харченко, “Концепція побудови сертифікаційної моделі якості програмних систем,” *Проблеми програмування*, no. 2-3, 275-281, 2006

Дослідження асинхронних методів сервісної взаємодії у веб-додатках

Микола Пікуляк, Станіслав Домбровський

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація—Проведено порівняльний аналіз протоколів асинхронної взаємодії сервісів, що дозволило виділити їхні позитивні та негативні сторони. Визначено напрямок покращення сервісної взаємодії у веб-додатках, направлений на розробку методу з використанням протоколу gRPC на основі написаних protobuf сервісів та генерації Java класів. Виконано тестування розробленого методу у веб-додатку для анонімних дискусій, який надає можливість спілкуватись користувачам за допомогою чату. Отримано результати дослідження, які підтвердили ефективність застосування запропонованого методу, оскільки його використання дозволило при навантаженні зменшити як час створення об'єктів так і час отримання значення атрибута в додатку. Також покращення дозволило збільшити пропускну здатність при великій кількості паралельних підключень. Визначено перспективи подальших досліджень, направлених на розробку нових методів покращення для інших мов програмування, які підтримують генерацію gRPC сервісів.

Ключові слова— *Протоколи сервісної взаємодії, Метод покращення асинхронної взаємодії, Веб-додатки, Клієнтські класи, Серверні класи.*

I. ВСТУП

У сучасному світі різного роду веб-додатки стають все популярнішими та охоплюють все ширші предметні області застосування. У зв'язку з цим виникає необхідність аналізу процесу взаємодії між клієнтом та сервером, оскільки це є критично важливим для успішної роботи будь якого додатку. Також зростаюча кількість веб-додатків стикається з проблемою великої кількості запитів, які відправляються на сервер.

Тому сьогодні проводиться багато розробок, метою яких є використання асинхронних методів сервісної взаємодії, які дозволяють виконувати більш ефективні запити до сервера, використовуючи менше ресурсів та зменшуючи тривалість обробки.

Асинхронність є концепцією програмування, яка означає, що дії можуть виконуватися паралельно та незалежно одна від одної. У контексті веб-додатків, асинхронність полягає в тому, що запити клієнта можуть бути оброблені паралельно з іншими запитами, що дозволяє зменшити час очікування відповіді та покращити продуктивність веб-додатку [1].

Дослідження асинхронних методів сервісної взаємодії також дозволяє розглянути питання безпеки та стійкості веб-додатків. Використання асинхронних методів також може допомогти забезпечити розширюваність та простоту масштабування веб-додатка, що забезпечує йому роботу з більшим обсягом даних та більшою кількістю користувачів.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Стосовно синхронної та асинхронної обробки запитів у веб-додатках на даний час є достатньо багато досліджень та розроблено ряд методів, направлених на використання різного роду протоколів з метою покращення взаємодії між користувачами та сервісами [2].

Зокрема, в роботі [1] розглянуто можливості забезпечення надійності та безпеки застосування асинхронної технології, в [3] – проаналізовано можливість використання кешування результатів веб-сервісів для зменшення часу відповіді.

Для забезпечення асинхронної взаємодії сервісів сьогодні використовуються такі відомі протоколи:

- gRPC: відкритий протокол, розроблений компанією Google. Він базується на протоколі HTTP/2 та підтримує різні мови програмування. gRPC використовує Protobuf для серіалізації даних та надає підтримку асинхронного програмування за допомогою потоків [4].

- WebSocket: протокол відкритого коду для двосторонньої комунікації між клієнтом та сервером в режимі реального часу [5]. Він може використовуватись для передачі різних типів даних, включаючи текст, бінарні дані, JSON, XML тощо. WebSocket також підтримує асинхронну взаємодію між клієнтом та сервером.

- MQTT: легкий протокол для IoT-пристроїв та мережі, який підтримує асинхронну взаємодію. MQTT використовує модель "підписки/публікації" для передачі повідомлень між клієнтами та серверами [6].

- AMQP: протокол для асинхронної взаємодії, який підтримує розподілену обробку повідомлень та різні типи повідомлень. AMQP використовується для побудови великих систем з обробкою повідомлень та мікросервісів [7].

- Apache Kafka: розподілена платформа для обробки повідомлень, яка підтримує асинхронну взаємодію та забезпечує збереження повідомлень на сервері. Вона використовується для обробки великих потоків даних та побудови потужних систем з обробки повідомлень.

Вибір протоколу або фреймворку для асинхронної взаємодії сервісів залежить від конкретних потреб та напрямку проекту, оскільки описані протоколи мають свої переваги та недоліки. Наприклад, gRPC може бути корисним для веб-додатків, які мають велику кількість мікросервісів та потребують швидкої та ефективної взаємодії між ними. Він використовує бінарний формат обміну даними Protobuf, що зменшує розмір даних та збільшує продуктивність передачі.

WebSocket, з іншого боку, може бути корисним для веб-додатків, які потребують надійного та тривалого зв'язку між клієнтом та сервером, з можливістю обміну даними в режимі реального часу.

Перевагою AMQP є забезпечення захисту повідомлень за допомогою шифрування та автентифікації, що важливо для програм, де безпека даних є ключовою. Kafka забезпечує високий рівень надійності, завдяки реплікації та розподіленій архітектурі.

Тому подальші дослідження асинхронних методів сервісної взаємодії у веб-додатках, направлених на підвищення продуктивності та ефективності їх використання, зменшення часу очікування та покращення користувацького досвіду є актуальною задачею, яка потребує подальших як теоретичних розробок так і практичних вдосконалень.

III. МЕТОДИ

В основі дослідження є розробка методу для покращення асинхронної взаємодії веб-сервісів з використанням протоколу gRPC разом з Protobuf, що дає можливість підвищити продуктивність та ефективність у веб-додатках за рахунок покращення генерації Java класів на основі написаних protobuf сервісів.

Для аналізу та покращення протоколу gRPC було розроблено веб-додаток, який дозволяє забезпечити асинхронну передачу даних від клієнта до сервера та між сервісами додатку при одночасному використанні. За основу для покращення було обрано додаток для анонімних дискусій, спілкування між користувачами в якому організовано за допомогою чату.

Запропонований клієнт-серверний додаток побудований на основі фреймворку SpringBoot та протоколів gRPC та Protobuf, додатково використано протокол WebSocket для отримання актуальної інформації для чатів. Така архітектура дозволяє з легкістю контактувати з веб-фронтом за допомогою REST підходу та слідкувати за правильною

роботою додатку. З іншої сторони веб-сервіс контактує з допоміжними чат сервісом та логін сервісом за допомогою протоколу gRPC та отримує відповідну інформацію (рис. 1):

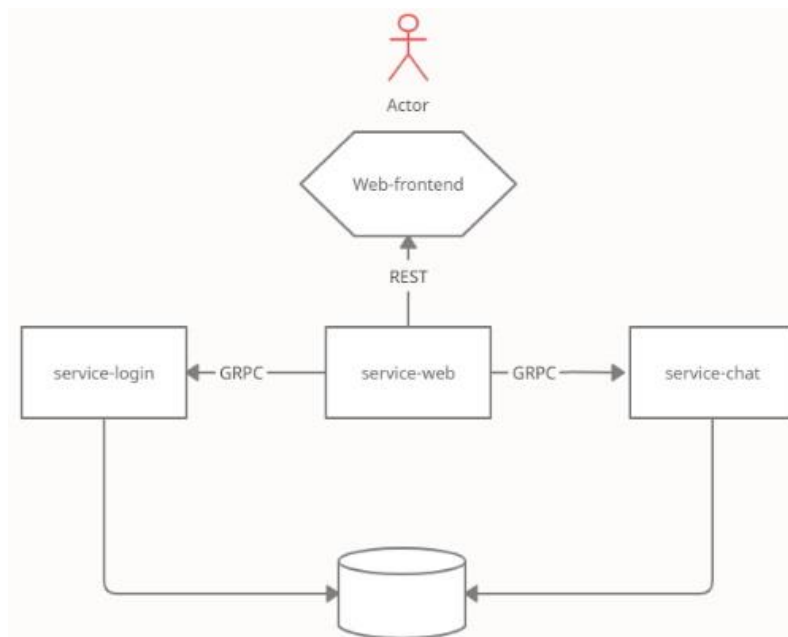


Рисунок 1 – Архітектура веб-додатку

Для зберігання інформації про чати, клієнтів та питання було використано базу даних PostgreSQL та створено відповідні таблиці chats, chat_clients, chat_questions. Перевагою використання PostgreSQL є те, що доступ з програм до бази даних здійснюється за допомогою спеціального процесу, під час якого клієнтські програми не можуть отримувати самостійний доступ до даних навіть у тому випадку, якщо вони функціонують на тому самому ПК, на якому здійснюється серверний процес.

Однією з ключових особливостей gRPC є можливість генерувати клієнтський код різними мовами, такими як C++, Java, Python [8].

Пересилка повідомлень protobuf за допомогою протоколу gRPC в програмному вигляді відбувається шляхом використання автоматично згенерованих клієнтських та серверних класів, створених на основі оголошень сервісу та повідомлень у файлах .proto.

Повідомлення Protocol Buffers (protobuf) представляються у бінарному, компактному форматі. Кожен тип даних та повідомлення мають свій внутрішній байтовий формат для серіалізації та десеріалізації даних (табл.1).

Таблиця 1 – Групи змінних у байтовому форматі

Тип	Значення	Використовується для
0	Varint	int32, int64, uint32, uint64, sint32, sint64, bool, enum
1	64-bit	fixed64, sfixed64, double
2	Length-delimited	string, bytes, embedded messages, packed repeated fields
3	Start group	groups (застарілий тип)
4	End group	groups (застарілий тип)
5	32-bit	fixed32, sfixed32, float

У protobuf кожен елемент даних має свій ключ (ціле число), яке ідентифікує елемент і тип даних, який визначає, які дані можуть бути включені в цей елемент. Типи даних

включають цілі числа, змінні довжини, фіксовані довжини, числа з рухомою комою, булеві значення і рядки.

Однією з центральних складових gRPC є генерація коду для клієнтських та серверних додатків. Цей код генерується компілятором protoc з файлів .proto.

На перший погляд, gRPC виглядає як ідеальне рішення, але є певні проблеми, не прямо пов'язані з gRPC або Protobuf, але з виборами, зробленими конкретно в генерованому коді.

Було вирішено покращити генерацію коду для структур з пакету com.google.protobuf, оскільки функціонал можна залишити незмінним, але спростити структуру класів.

Для прикладу розберемо поле "question" типу "com.google.protobuf.StringValue". Генерований код надає можливість визначити, чи поле містить певне значення, отримати значення у вигляді структури "com.google.protobuf.StringValue" та помістити значення того ж типу.

```
boolean hasQuestion();
com.google.protobuf.StringValue getQuestion();
public Builder setQuestion(com.google.protobuf.StringValue value);
```

Саме методи get та set є не завжди зручними і оптимальними, оскільки для полів типу string отримувати складну структуру, в якій міститься потрібне значення, є зайвим.

Теж саме стосовно методу set; для того щоб помістити значення, потрібно викликати білдер класу StringValue та виконати додаткові операції. Для даного прикладу створення об'єкта типу ChatQuestions з полем Question буде мати такий вигляд:

```
ChatQuestions.newBuilder()
    .setQuestion(StringValue.newBuilder())
    .setValue("value")
    .build()
    .build();
```

Під даним покращенням мається на увазі змінити генерацію коду Java класів опису proto структур для комплексних структур типу ...Value, з додаванням допоміжних методів get...Value та set...Value, які б приймали та віддавали прості типи без надбудов.

Таке покращення направлене на зменшення кількості операцій під час створення та отримання об'єктів, які містять ...Value поля для збільшення перформенсу, а також для спрощення написання програм – оптимізації доступу до даних. Покращені класи нададуть більш швидкий та зручний доступ до даних, що сприяє підвищенню продуктивності додатку, а також покращення генерації коду призведе до зменшення споживання пам'яті та обчислювальних ресурсах.

IV. РЕЗУЛЬТАТИ

Для дослідження було обрано навантажувальне та об'ємне тестування.

Навантажувальне тестування виконується для оцінки поведінки додатку із заданим очікуваним навантаженням. Наприклад, кількістю користувачів, які використовують систему одночасно. Об'ємне тестування передбачає зміну розміру даних, які використовуються для оцінки поведінки системи.

В результаті проведеного дослідження із використанням звичайного та покращеного методів було протестовано час створення об'єктів з різною кількістю користувачів та кількістю даних.

Для вимірювання показників при навантаженні проекту до покращення та після було використано сервіс Guava від Google, оскільки він є простим інструментом для вимірювання часу та корисним для використання в будь-якому місці програми. Цей інструмент є потокобезпечним, тому можна використовувати кілька секундів паралельно без впливу

2023 International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

на продуктивність. Кожне завдання, яке потрібно виміряти за допомогою секундоміра, зберігається в Hashtable.

Під час порівняння було проведено аналіз часу побудови protobuf об'єктів з та без покращення в розмірі 500000 об'єктів (рис. 2):

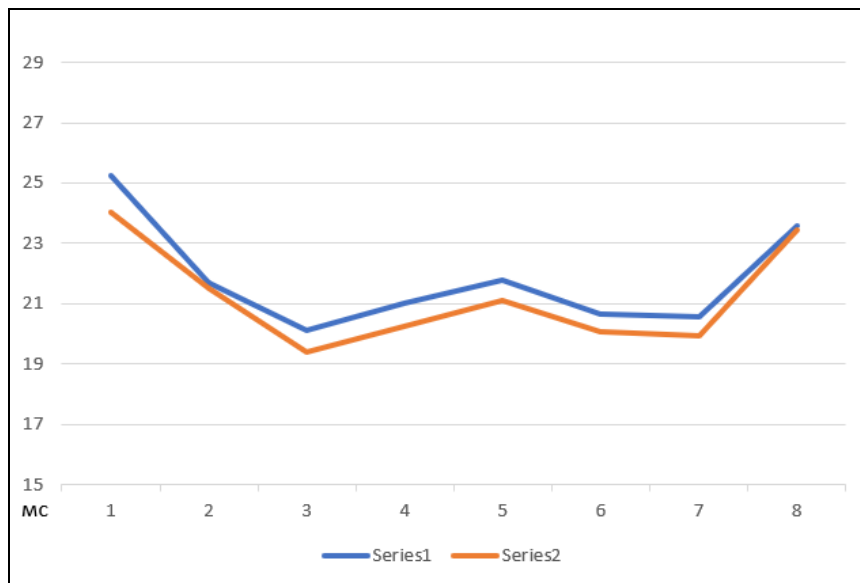


Рисунок 2 – Порівняння продуктивності створення об'єктів

Створено діаграму для демонстрації результатів порівняння. При восьми перевірках результати мали певну розбіжність та в середньому результат показує, що покращення призвело до зменшення часу створення об'єктів в додатку при навантаженні на 2.8%.

Для наступного порівняння було проведено аналіз часу отримання значення атрибута з уже створеного protobuf об'єкта з та без покращення в розмірі 500000 разів (рис. 3):

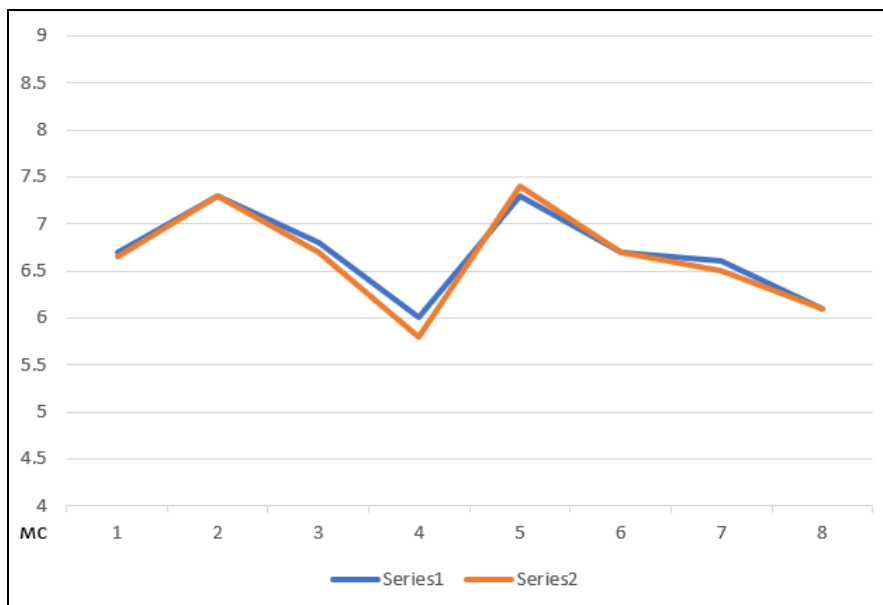


Рисунок 3 – Порівняння продуктивності отримання значення атрибута

Створено діаграму для демонстрації результатів порівняння. Кількість перевірок така ж сама, результати також мали певну розбіжність. Проте в середньому результат показує, що

покращення призвело до зменшення часу отримання значення атрибута в додатку при навантаженні на 0.4%

Також для порівняння результатів передачі даних від клієнта до сервера протоколом gRPC було вирішено перевірити на прикладі двох тестових сценаріїв з різною кількістю користувачів. В першому сценарії брались до уваги 10 користувачів, в другому 100. Кожний користувач був представлений як окремий потік Java, але дані передавались через однаковий серверний стрім та розмір повідомлень був однаковий для кожного з них.

Щоб вивчити залежність продуктивності протоколу від кількості повідомлень було вирішено розділити тестові дані на два різні набори даних. Таким чином, результатом першого тесту було отримання від сервера малої кількості повідомлень в розмірі 100 штук та 1000 штук в другому тесті.

Кожен набір даних перевірявся окремо в різному тестовому сценарії (табл. 2-3).

Таблиця 2 – Результати порівняння для малої кількості паралельних підключень (10)

	100 повідомлень		1000 повідомлень	
	Без покращення	З покращенням	Без покращення	З покращенням
Середній час передачі, мс	0.029	0.028	1.27	1.24

Згідно таблиці 2 можна стверджувати, що покращення дає певне збільшення пропускної здатності. При малій кількості отриманих повідомлень покращення потребує на 3.5% менше часу та на 2.4% менше часу при більшій кількості надісланих даних.

Таблиця 3 – Результати порівняння для великої кількості паралельних підключень (100)

	100 повідомлень		1000 повідомлень	
	Без покращення	З покращенням	Без покращення	З покращенням
Середній час передачі, мс	0.66	0.64	9.5	9.3

Відповідно до результатів, представлених в таблиці 3 видно, що покращення дозволяє збільшити пропускну здатність також при великій кількості паралельних підключень. При малій кількості отриманих повідомлень покращення потребує на 3% менше часу та на 2.1% менше часу ніж при більшій кількості надісланих даних.

V. ОБГОВОРЕННЯ

Отримані результати демонструють важливість використання асинхронних методів сервісної взаємодії у веб-додатках, оскільки дають можливість оптимізувати використання системних ресурсів, підвищити продуктивність та ефективність під час передачі даних від клієнта до сервера та між сервісами додатку. Застосування запропонованого покращеного методу з використанням генерації Java класів дозволяє отримати більш швидкий та зручний доступ до даних, що сприяє зменшенню споживання пам'яті та збільшенню пропускної здатності при великій кількості паралельних підключень.

VI. ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ

Подальшим напрямком дослідження буде розробка методів, які дозволять оцінити та покращити часові параметри сервісної взаємодії під час навантажувального та об'ємного тестування для інших мов програмування, які підтримують генерацію gRPC протоколів.

VII. ВИСНОВКИ

В роботі виконано аналіз протоколів асинхронної взаємодії, які використовуються у веб-додатках з метою підвищення ефективності їх використання.

Запропоновано метод для покращення асинхронної взаємодії веб-сервісів з використанням протоколу gRPC на основі написаних protobuf сервісів та генерації Java класів, що дозволило надати більш швидкий та зручний доступ до даних та підвищити продуктивність додатку.

Тестування розробленого методу було виконано за допомогою розробленого веб-додатку для анонімних дискусій, який надає можливість спілкуватись користувачам за допомогою чату. Отримані результати дослідження, проведеного у сервісі Guava із використанням звичайного та покращеного методів показали, що використання розробленого покращеного методу дозволило зменшити час створення об'єктів в додатку при навантаженні на 2.8%, а час отримання значення атрибута в додатку при навантаженні зменшити на 0.4%. Також покращення дозволило збільшити пропускну здатність при великій кількості паралельних підключень.

ЛІТЕРАТУРА

- [1] С. Крузе, О. Союз, "Synchronous and Asynchronous Web Service Interaction", *International Journal of Web Services Research*. 2013.
- [2] М. Дутчак, М. Пікуляк, "Оцінка якості методів та програмних засобів інтелектуальних адаптивних освітніх вебсистем самонавчання," *2022 International Conference on Innovative Solutions in Software Engineering (ICISSE)*, Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, Nov. 29-30, 2022, pp. 266-270.
- [3] S. Varghese. Building High Performance APIs In Go Using gRPC And Protocol Buffers [Електронний ресурс]. Режим доступу: <https://shijuvar.medium.com/building-high-performance-apis-in-go-using-grpc-and-protocol-buffers-2eda5b80771b>. Дата звернення: Жовт. 18, 2023.
- [4] gRPC – A high performance, open source universal RPC framework. [Електронний ресурс]. Режим доступу: <https://grpc.io/>. Дата звернення: Вер. 16, 2023.
- [5] WebSocket. [Електронний ресурс]. Режим доступу: <https://uk.wikipedia.org/wiki/WebSocket>. Дата звернення: Жовт. 26, 2023.
- [6] Чому MQTT є важливою частиною IoT? [Електронний ресурс]. Режим доступу: <https://fiberroad.com/uk/why-is-mqtt-an-important-part-of-iot/>. Дата звернення: Жовт. 21, 2023.
- [7] AMQP is the Internet Protocol for Business Messaging. [Електронний ресурс]. Режим доступу: <https://www.amqp.org/about/what>. Дата звернення: Жовт. 28, 2023.
- [8] S. G. Du, J. W Lee, K. Kim "Proposal of GRPC as a new northbound API for application layer communication efficiency in SDN, " *IMCOM '18: the 12th international conference on ubiquitous information management and communication, Langkawi Malaysia*. New York, NY, USA, 2018. DOI: <https://doi.org/10.1145/3164541.3164563>.

СИСТЕМНИЙ АНАЛІЗ АКВАПОННИХ СИСТЕМ

Роман Залозний, Наталія Заєць

*Кафедра автоматики та робототехнічних систем імені академіка І.І.Мартиненка
Національний університет біоресурсів і природокористування України
Київ, Україна*

Аквапоніка - це інноваційна, розумна та стійка сільськогосподарська технологія, яка інтегрує аквакультуру (розведення риби) з гідропонікою у симбіотичному вирощуванні овочевих культур. Особливість аквапонних систем полягає у тому, що вони є замкненими циклами з великою кількістю змінних. Тому для ідентифікації ключових елементів та взаємозв'язків було розроблено графову модель. Використання графової моделі забезпечує можливість побудови матриці шляхів, що в свою чергу дає можливість робити висновки про складність системи з точки зору системного аналізу.

Ключові слова— Аквапонна система, Системний аналіз, Графова модель, Матриця шляхів

I. ВСТУП

Аквапоніка стає все більш популярною та важливою в галузі сталого розвитку сільського господарства та виробництва продуктів харчування. При дослідженні та розробці аквапонних систем важливу роль відіграє системний аналіз. Оскільки аквапоніка об'єднує аквакультуру, гідропоніку та циркуляцію води, що створює складну взаємодію між живими організмами, рослинами, водою, добривами та іншими елементами системи. Системний аналіз дозволяє виявити взаємозв'язки та залежності між елементами системи, виявити елементи системи, які можуть бути оптимізовані, щоб підвищити продуктивність та ефективність технологічного процесу.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Існуючі дослідження в галузі аквакультури, гідропоніки та аквапоніки, як єдиної інтегрованої системи, показують, що розробка математичних моделей та проведення системного аналізу є початковим кроком у напрямку створення надійної та ефективної системи управління технологічним процесом.

Більшість існуючих досліджень у цій області можна розділити на кілька категорій. Перша категорія статей присвячена системам управління, включаючи використання Інтернету речей. Зокрема, у роботах [3, 8, 15] впроваджено системи моніторингу технологічних параметрів у реальному часі, що призвело до підвищення ефективності аквапоніки та зменшення витрат.

Інша категорія досліджень присвячена оптимізації параметрів технологічних процесів, при цьому головними параметрами в аквапонних системах є гідрохімічні параметри. Роботи [4, 10] розглядають основні технологічні аспекти аквапоніки і разом з використанням технології Інтернету речей сприяють якісному виробництву. Дані статті також надають загальні діапазони для підтримки параметрів, таких як рівень рН, температура, освітленість, лужність та інші ключові параметри для більшості рослин і гідробіонтів.

Підтримка оптимальних параметрів для рослин та риб також має вплив на їх здоров'я. У роботах [7, 9] представлені результати досліджень впливу добрив на життя гідробіонтів.

III. МЕТОДИ

Системний аналіз дозволяє визначити основні взаємозв'язки у технологічному процесі. А використання графової теорії є ефективним інструментом для визначення складності системи.

Графові моделі дозволяють зобразити складну структуру системи у зрозумілій та зручній формі. Це дозволяє легше розуміти взаємозв'язки та взаємодію об'єктів. Застосовуючи алгоритми графового аналізу, можна виявити ключові елементи, знаходити шляхи сполучення, визначати вплив об'єктів на систему тощо. У свою чергу це дає змогу прогнозувати поведінку системи та виконувати оптимізацію параметрів.

Створення графової моделі можливе лише при ретельному дослідженні структури аквапонних систем. Загалом, у аквапонних системах можна виділити три основні групи підсистеми, навколо яких і здійснюється функціонування технологічного процесу. Це: система аквакультури, система гідропоніки та система керування.

IV. РЕЗУЛЬТАТИ

Для системного аналізу аквапоніки доцільно використовувати теорію графів, як один з методів математичного опису зв'язків між підсистемами. Графові моделі поєднують в собі характеристики графічного та множинного представлення, а також формують допоміжні матриці, які зручно використовувати для аналізу, синтезу та управління. Аквапонний комплекс - це складна система, що включає в себе такі елементи:

1) Датчик рівня освітленості; 2) Датчик кисню; 3) Датчик вологості повітря; 4) Датчик рівня води рослини; 5) Датчик рівня води риби; 6) Датчик температури води; 7) Датчик температури повітря; 8) Датчик провідності води; 9) Датчик рН; 10) Система штучного освітлення рослин; 11) Аератор; 12) Система подачі корму; 13) Насос прямий; 14) Фільтр відходів; 15) Біофільтр; 16) Насос зворотній; 17) Насос зовнішній; 18) Підігрів води; 19) Підігрів повітря; 20) Система подачі додаткових мінералів; 21) Зволожувач повітря; 22) Резервуар з рибами; 23) Система автоматичного керування технологічним процесом; 24) Резервуар з рослинами.

Кожна підсистема, яка виділена, є важливою складовою аквапоніки. Аналіз цих підсистем сприятиме отриманню більш повного розуміння самого процесу та можливості його регулювання. Шляхом проведення технологічного та структурного аналізу був складений граф структури аквапонної системи (рис. 1), де функціональні елементи є вершинами, а зв'язки між ними позначаються дугами.

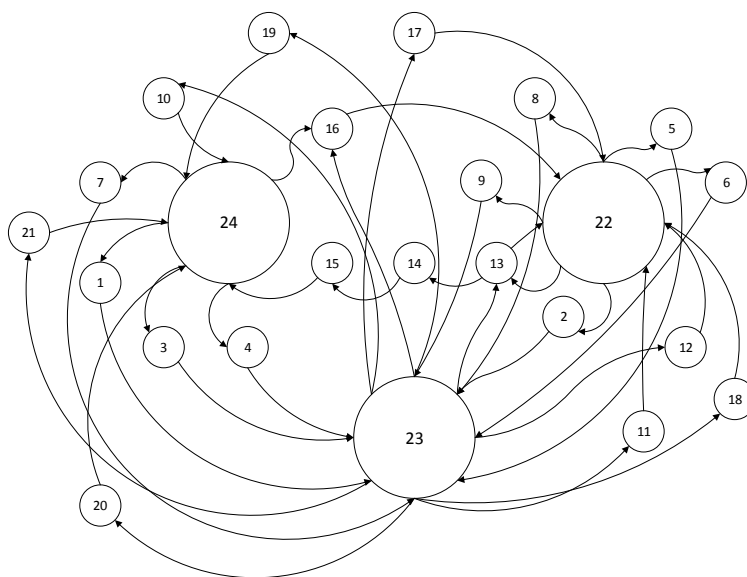


Рисунок 2 – Графова модель структури аквапонної системи

V. ОБГОВОРЕННЯ

В структурі аквапонної системи виділяється двадцять дві окремі підсистеми та три комплекси: 22 - резервуар з рибами, 23 - система керування і 24 - резервуар з рослинами. Окремі підсистеми, як показано на Рис.1, є незалежними компонентами аквапоніки. Отже, для забезпечення їхньої стабільності потрібне окреме керування. Це ускладнює сам процес, але забезпечує більшу стабільність системи і спрощує впровадження системи керування.

Після аналізу графової моделі структури аквапонної системи, матриці шляхів та допоміжної матриці стає очевидним, що аквапонна система повинна враховувати вплив окремих компонентів підсистеми на загальний режим роботи комплексу, включаючи взаємодію між резервуаром з рибами та резервуаром з рослинами, а також функціонування компонентів, які є частиною обох підсистем і системи керування. З урахуванням того, що 22 та 24 (згідно з Рис. 1) можуть функціонувати незалежно, система керування має здійснювати компенсацію взаємних перешкод та окремих флуктуацій.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Подальші дослідження у галузі аквапоніки сприятимуть розробці системи управління, що дозволить досягти оптимальних управлінських параметрів та встановити баланс між впливами всіх підсистем на загальний режим функціонування. Системний аналіз аквапонних систем допомагає сформулювати ключові етапи для забезпечення ефективності виробництва: ідентифікація складових системи, поетапну оцінку витрат енергії, оцінку продуктивності, оптимізацію, моніторинг та управління системами, а також впровадження системного підходу..

VII. ВИСНОВКИ

Аналіз дозволяє прийти до висновку, що розглядуваний об'єкт є складною системою, що складається з 22 окремих підсистем і трьох комплексів: комплексу із резервуаром для риб, комплексу із резервуаром для рослин і системи керування.

Проведення аналізу графу та матриць шляхів дозволяє розробити і впровадити практичні аспекти управління, зокрема: Спростити процес управління та забезпечити більший контроль над кожною складовою системи. Здійснювати аналіз діагностики та виявлення можливих несправностей окремо для цих підсистем та комплексів, що допомагає вчасно виявляти проблеми та уникати їх поширення на інші частини системи. Оптимізувати ресурси для забезпечення оптимального функціонування всього комплексу.

ЛІТЕРАТУРА

1. Ahmad Zatznika Purwalaksana, Indra Hartarto Tambunan, Andi Ray Hutauruk. (2022) Automation and Monitoring for Aquaponic System Based on NodeMcu, Journal of Technical Engineering: Piston Vol. 5/ No. 2, 72-82
2. Ariyawan Sunardi , Fahmi Islami Suud, N Woro Agus, Indra Gunawan. (2018) IoT Application on Aquaponics System Energy Optimization, The 2nd International Conference on Energy Sciences (ICES), Bandung, Indonesia, Journal of Physics: Conference Series, Volume 1772.
3. Divas Karimanzira, Thomas Rauschenbach. (2021) An intelligent management system for aquaponics. Взято з: https://www.researchgate.net/publication/350632213_An_intelligent_management_system_for_aquaponics
4. Maritel Dawa, Maria Rodelyn Tibon, Samson Mahidlawon Lausa.(2022) Internet of things (IoT) based aquaponics management system adaptive to climate change. Взято з: https://www.researchgate.net/publication/364792727_Internet_of_things_IoT_based_aquaponics_management_system_adaptive_to_climate_change

5. Milena Krastanova, Ivo Sirakov, Sofiya Ivanova, Dobry Yarkov. (2022) Aquaponic systems: biological and technological parameters. Взято з: https://www.researchgate.net/publication/360427130_Aquaponic_systems_biological_and_technological_parameters
6. Mohamed Farag Taha, Gamal ElMasry, Mostafa Gouda, Lei Zhou, Ning Liang, Alwaseela Abdalla, David Rousseau, Zhengjun Qiu. (2022) Recent Advances of Smart Systems and Internet of Things (IoT) for Aquaponics Automation: A Comprehensive Overview. Взято з: https://www.researchgate.net/publication/362425402_Recent_Advances_of_Smart_Systems_and_Internet_of_Things_IoT_for_Aquaponics_Automation_A_Comprehensive_Overview
7. Zahari Abu Bakar, Zairil Nor, Kamaru ADZHA Kadiran, Mohamad Farid Misnan. (2022) Smart Plant Monitoring System Using Aquaponics Production Technological with Arduino Development Environment (IDE) and SMS Alert: A Prototype. Взято з: https://www.researchgate.net/publication/365840530_Smart_Plant_Monitoring_System_Using_Aquaponics_Production_Technological_with_Arduino_Development_Environment_IDE_and_SMS_Alert_A_Prototype

Розробка методології автоматизованого тестування мобільних застосунків

Остражнов Владислав

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація — Робота присвячена розробці системи автоматизованого тестування мобільних застосунків. У контексті швидкого росту мобільних технологій, забезпечення якості програмного забезпечення для мобільних платформ стає дедалі важливішою задачею. Розглядаються сучасні методи та підходи до автоматизованого тестування, а також представлена пропонувана система, що дозволяє ефективно виконувати тестування на різних мобільних пристроях та платформах.

Ключові слова — *Інженерія програмного забезпечення, Тестування, Автоматизоване тестування, Мобільні застосунки, Якість програмного забезпечення.*

I. ВСТУП

На сьогоднішній день мобільний телефон став невід’ємною частиною великої кількості людей завдяки мобільним додаткам, він допомагає забезпечити організоване життя. Про це свідчить наявність програм для контактів, відповідних проектів та подій, особистої інформації у мобільних телефонах. Цей набір інформації повністю зберігається на наших телефонах і може допомогти нам планувати життя, тим самим полегшуючи належне управління часом. Існує багато програм, такі як погода, календар зустрічей, списки справ та всі типи додатків для сповіщень, які можна запрограмувати на нашу індивідуальні потреби та вимоги, роблячи життя комфортнішим, простішим та продуктивнішим [1,2]. Саме тому актуальність розробки додатків залишається по цей час. На фоні великої кількості додатків тільки малий відсоток покритий тестами, що, як наслідок приводить до появи недосконалих, незручних або взагалі продуктів, що не можна використовувати. Для покращення якості продукту та виявлення усіх багів потрібно звикористовувати автоматизоване тестування, яке для більшості розробників є надто часозатратним та дорогим.

Тобто актуальним є розробка систем автоматизованого тестування, які дозволяють зменшити час створення та виконання самих тестів, що в результаті дозволить охопити значну кількість додатків [3].

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

В результаті проведеного аналізу публікацій та підходів до автоматизованого тестування, було знайдено методу, яка включає в себе використання Appium драйвера, фреймворка TestNG та мови програмування Java. Особливостями цієї методу є написання тестів послідовно зарахунок прописування напряму ідентифікаторів у тестовому класі, даний спосіб допомагає у написанні декількох швидких тестів для перевірки деякого функціоналу, але це надто довго при написанні більше п’яти тестів. Також одним з мінусів даного методу це створення цих самих автотестів без завчасно продумання та написання тестового плану, що часто може призвести до малого відсотка покриття мобільного застосунка тестами.

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

Останнім мінусом у знайденій методиці являється запуск тестів у послідовному режимі, замість паралельного, що суттєво сповільнює виконання автоматизованого тестування. Відповідно актуальність впровадження методології в автоматизованому тестуванні, яка б зменшувала час на написання автотестів та їх виконання, а також була абсолютно зрозуміла для спеціалістів у цій сфері, є дуже високою.

III. МЕТОДИ

Основна мета пропонованої методики полягає у скороченні часу на написання автоматизованих тестів. Основним критерієм, що свідчить про ефективність проведеного дослідження, є затрати часу написання автотестів згідно стандартної методики та часу, за який можливо написати ті самі тести з використанням пропонованої методології. Одним із ключових елементів методології, що скоротить час автоматизованого тестування як такого, полягає у запуску тестів паралельно.

IV. РЕЗУЛЬТАТИ

Першим кроком було групування всіх сторінок застосунку по класах, що дає чітке розуміння розташування всіх елементів і сторінок додатку, та дозволяє прописати локатори тільки один раз для елементів та пізніше їх перевикористовувати бажану кількість разів.

Другим кроком після групування класів, проведено попередню ініціалізацію та створення необхідних методів, що зменшує кількість кроків до написання безпосередньо у тестовому класі, їх повторення у разі перевикористання та неактуальність пошуку ідентифікаторів у тестовому класі, так як усі ідентифікатори на потрібних сторінках вже були прописані та знайдені.

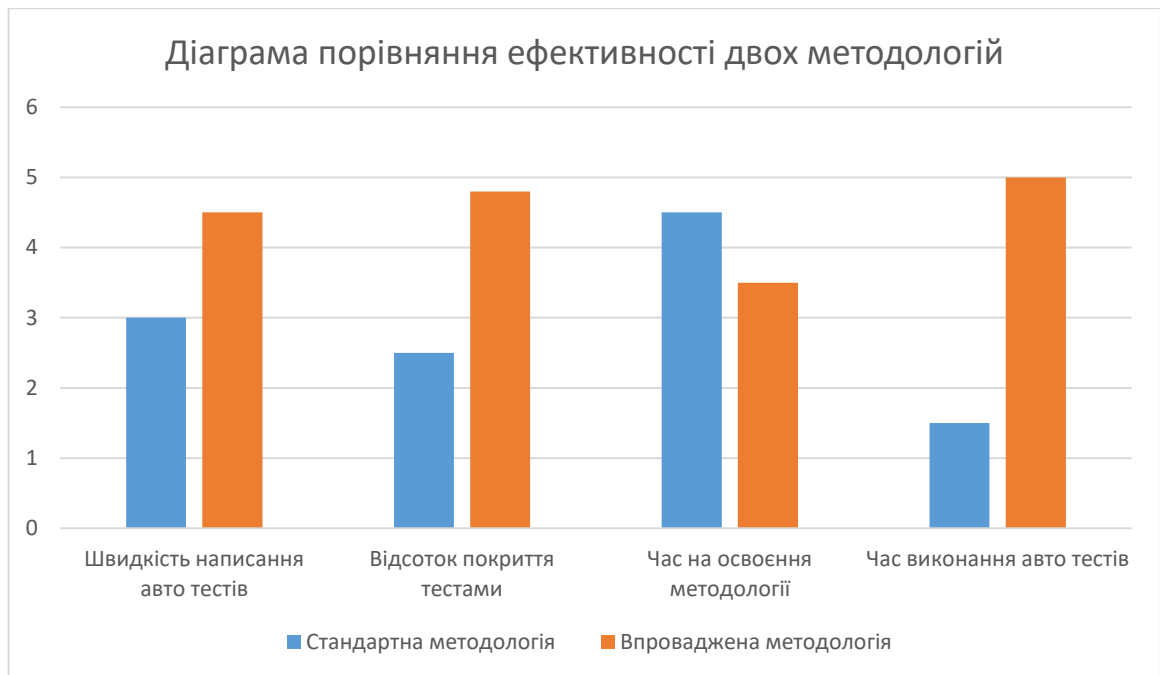
Третім кроком був завчасно написаний вручну тестовий план та тестові сценарії, які структурують написаний тестовий клас, його порядок, відсоток покриття застосунку тестами з різними кейсами, як негативними, так і позитивними що відображає чіткий та зрозумілий порядок тестів.

Четвертим та фінальним кроком було впроваджено паралельний запуск тестів, що дозволило в кілька разів збільшити швидкість виконання тестів порівняно із послідовним методом запуску.

V. ОБГОВОРЕННЯ

Запропонована модель, згідно отриманих результатів, дозволяє розробляти та виконувати автотести швидше у порівнянні із традиційними підходами. Як наслідок, це дозволить охопити більший відсоток мобільних застосунків, покращити їх якість, бо дозволить змінити ставлення розробників до систем автоматизованого тестування як до необхідної та невідомої частини розробки мобільного застосунку.

Згідно проведених досліджень та використання методології запропонованої в інтернеті та розробленої методології, було виконано написання авто тестів та їх запуск з одним і тим самим сценарієм, відповідно до цих результатів було створено діаграму 5.1, де оцінка 5 є найвищою та 0 відповідно найнижчою.



Діаграма 5.1 – Порівняння ефективності двох методологій

Згідно представленої діаграми можна зробити висновок що методологія дає великий приріст в ефективності для кожної з метрик, окрім часу на освоєння методології, адже все таки вона є кращою за стандартну і трохи складніша відповідно, але згідно оцінки втрати не є суттєвими. Також необхідно звернути увагу на пункт “Час виконання авто тестів”, у впровадженій метриці де використовується паралельний запуск авто тестів було використано власний комп’ютер, відповідно при використанні потужнішого комп’ютера швидкість виконання авто тестів може зрости порівняно з послідовним запуском тестів.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Наступним кроком для покращення запропонованої методології може бути повна реалізація утиліти, яка включала би в себе усі властивості Appium server, Appium inspector, Browser Stack та можливість відображення ідентифікаторів елементів при відсутності id.

VII. ВИСНОВКИ

Проведені дослідження дозволили реалізувати методологію, швидкість написання та виконання автоматизованого тестування завдяки 4 запропонованим крокам, значно зростає. Першим кроком було чітке групування класів та завчасне знаходження ідентифікаторів елементів, другим - ініціалізація класів та створення методів для подальших автотестів, третім - розробка чітких сценаріїв згідно яких пишуться автотести та четвертим - паралельний запуск тестів.

ЛІТЕРАТУРА

- [1] Ручне та автоматизоване тестування. URL: <https://qalight.ua/baza-znaniy/ruchne-ta-avtomatizovane-testuvannya/> (дата звернення 26.10.2023).
- [2] Автоматизоване тестування. URL: https://uk.wikipedia.org/wiki/Автоматизоване_тестування (дата звернення 26.10.2023).

- [3] Фундаментальний процес тестування. URL: <https://qalight.ua/baza-znaniy/fundamentalnij-protses-testuvannya/> (дата звернення 07.11.2023).

Modeling a method for generating a stream of secret keys in the form of permutation matrices for encryption-masking of video frames and studying its characteristics

Vladimir Krasilenko, Vladislav Podlubnyi, Diana Nikitovich

*Department of Computer Science and Economic Cybernetics
Vinnytsia National Agrarian University
Vinnytsia, Ukraine*

Abstract— The article considers a method of forming a stream of secret matrix keys in the form of permutation matrices. On the basis of consideration of the advantages of matrix models and cryptosystems for masking video frames, the urgent need to form a stream of secret matrix keys (MKs) is stressed. It is shown that, taking into account of crypto-transformations in matrix affine-permutation ciphers, a number of keys in the form of permutation matrices (MPs) are required for the successful use of the latter in frame masking tasks. To solve this problem the article proposes a approach of generating a series of MKs (MPs). The method is based on the use of a series of crypto-transformations of the base key using affine encryption while changing the keys of this cipher in accordance with the generated random sequence. Functionality and advantages of the proposed method are demonstrated by model experiments in the Mathcad, screenshots from the created modules. The properties of a set of generated MKs (MPs) were investigated using mutual correlation and equivalence normalized functions. Adequacy and stability of the method were confirmed. The advantage of the method is the focus on parallel processing, ease of adaptation to different formats of images, isomorphism of visualization of keys.

Keywords— *Cryptography transformations, Secret matrix key, Permutation matrix, Keys stream, Matrix model, Encryption-decryption, Masking of video frames, Cipher, Spatial equivalence function.*

I. INTRODUCTION

A Review of the publication

Today, the most prevalent type of data used in various fields, including industry, science, technical development, and everyday life, is digital visual information in the form of images and video frames. A vast number of documents containing scientific, technical, educational, and other data include diverse images of objects, diagrams, schematics, drawings, and photographs, both halftone and color [1-6]. Cryptographic methods for protecting information must be used today in a significant number of information technologies and areas of their application. These areas include the Internet of Things (IoT), intelligent sensors and devices for collecting, processing and transmitting information, microcontrollers and devices for diagnostics and monitoring in medicine, military equipment, RFID tags, network information and communication technologies, geo-information systems for remote monitoring [4-7].

Taking into account the diversity and specificity of these areas, different requirements are placed on cryptographic algorithms, models and systems. But there are also general requirements, such as stability and sustainability. The latter in most known cryptosystems significantly depend on the quality of the generated keys, for example, the gamma in stream ciphers [4, 6-9]. Determining the randomness and characteristics of the generated sequences is one of the priority tasks. A good generation of a pseudo-random sequence is one for which it is impossible to predict the next values

without knowledge of the seed, knowing the entire history of previous values. There are known works in which the statistical properties of pseudorandom sequences generated by the corresponding stream cryptographic algorithms and their hardware and software implementations were studied. Independent testing by the authors of the properties of such algorithms and generators in identical universal conditions made it possible to obtain objective and independent results of a comparative analysis and substantiate the principles of creating a new stream cryptographic algorithm, which can be the basis of the national standard of Ukraine [8, 9]. This is explained by the fact that the study of the “Strumok” algorithm, developed by Ukrainian scientists, showed that its statistical and other characteristics correspond to the best algorithms in the world [8, 9]. And therefore, in our work, we will show how to build new pseudo-random streams based on such generators, but not simple bit or numerical values, but secret keys of a special format

Despite the diversity of such information and the formats in which it is presented, all these text-graphic documents (TGD) are represented as sets of page images or their fragments, which are stored and displayed using various devices, including computer monitors and displays.

And this will make it possible to process large data sets or whole streams of video frames at an accelerated pace, to solve new and more complex tasks. The main aspect of relevance is the improvement of the basic characteristics of transmission processes, protection against unauthorized access and information hiding in telecommunication systems based on matrix models, their new transformation procedures. One of the urgent issues is the study of the prospects for the application of matrix models and transformations in algorithms for compression, masking of images and video frames, which requires separate research. Solving this issue will significantly increase the security of transmission of digital visual information in telecommunication channels.

The constant increase in the bandwidth of information transmission channels and the speed of its processing in communication systems, in hardware and software accelerators and computer architectures is compensated, firstly, by the constant growth of the volume of both public and confidential visual digital information, and secondly, the increasing requirements for the dimensionality and resolution of such information. This leads to the necessity improving, and often, revising the foundations of building methods and means of transforming such information in objects of distributed systems.

The fundamental image processing operations include compression and protection against unauthorized access and the influence of natural and artificial factors, and for this purpose, matrix transformations such as orthogonal and affine transformations are used.

Today, researchers are showing significant interest in discrete matrix transformations of information, both in a general context and in the realm of cryptography, for information encryption and protection. Expanding the set of basic matrix operations allows for selecting the most suitable operation for a specific task, and modern processors and programmable logic contribute to the implementation of these operations at the hardware level, enabling more efficient processing of large volumes of data, TGD, and video streams.

However, the primary challenge lies in enhancing the transmission, protection, and concealment of information in telecommunication systems using matrix models, their operations, and transformations. One of the urgent problems is the research and assessment of the possibilities of using an extended family of matrix models and transformations in algorithms for compression, hiding and protection of visual data, taking into account their features and properties. This requires additional research. Solving this problem will help to significantly increase the security of transmission of digital visual information through telecommunication channels.

Development of a method of frame-by-frame masking matrix transformation of visual data to protect against unauthorized access when storing images or video files and transferring them in open communications is a very urgent task, which has already been studied at the level of crypto-transformation models of individual images or individual frames. However, for the direct or reverse crypto-transformation of the entire flow of frames or image matrices based on new matrix models, additional research is required.

The solving of such a task, as forming a series of frame-by-frame masking encryption-decryption keys is required.

Advantages of cryptographic transformations (CT) of textographic documents (TGD) with visas, signatures, images (I), tables, diagrams, etc., in cryptosystems of the matrix type (MT) [10, 11, 13, 14] based on algorithms and matrix-algebraic models (MAMs), including generalized matrix affine and affine-permutation ciphers were demonstrated in works [11-15].

Modifications of MAM were used in the creation of blind and other digital signatures [16], they allow checking the presence of distortions in cryptograms of black and white and color images, their integrity [11], creating block [13-14], multifunctional parametric models [13], multi-page [14] and investigate their stability characteristics. The basic operations of MAM are element-by-element multiplication, addition modulo matrices, and matrix permutation models (MP_M) with matrix multiplication procedures.

For security purposes technologies of cryptography, tools for CTs and protocols for the formation of keys and their exchange [15-17] are used, but only small part is devoted to methods oriented on matrix models (MMs) [10] and tools. In work [15] generalized algorithms for CTs, so-called matrix affine-permutation ciphers (MAPCs) based matrix affinity ciphers (MACs), were proposed. The results of simulation [9] of CTs used such MMs have shown their significant advantages: greater stability, increase in speed. In our works based on of MACs the algorithm for creating digital blind signature (DBS) is proposed.

To implement cryptographic transformations, it is necessary to produce byte matrices, permutation matrices, and image matrices representing characters, codes, and bytes through permutations. To achieve uniform alignment of histograms of image spectral components and increase the entropy of the image cryptogram using matrix-algebraic model transformations, the decomposition of R, G, B channels and their bit components is required [18, 19]. This necessitates the generation of numerous matrix keys and vectors. In other words, there is a need to develop a series of matrix transformations and keys to address these tasks.

B Problem Statement

Creation of generators of sequences of random numbers evenly distributed in a given interval is one of the main problems of developing information protection systems. Generating long random sequences is one of the important problems of classical cryptography [16, 17]. Pseudorandom sequence generators (PSGs) are widely used to solve this problem. Generators of pseudo-random sequences must meet certain conditions [16, 17, 20, 21]. Sequences obtained with their help should have a uniform distribution (or at least close to uniform). This means that the number of zeros in the generated binary sequence should be approximately equal to the number of ones contained in the sequence.

The gamma must be unpredictable, which means that it is impossible to predict the next bits of the sequence (gamma) by the previous segments of its bits, even if the type of the generator or the algorithm of its operation are known. To create an almost unpredictable gamut, it must, at a minimum have a very large period and a uniform law of distribution of bits or possible values of

2023 2nd International Conference on Innovative Solutions in Software Engineering
Ivano-Frankivsk, Ukraine, November 29-30, 2023

numbers or code words from the alphabet, in general, over the entire length of the gamut. In addition, the random values that make up the generated sequence must be statistically independent. This requirement means that there should be no correlation between individual bits and between groups of bits. These requirements are provided by the application of one-sided and computationally complex mathematical problems, for example, quadratic remainders modulo, the complexity of solving the problem of factorization of numbers. In order for the generator to be efficient, it must generate a very long sequence in the shortest possible time. For real-time systems, this requirement is particularly important. At the same time, for the use of generators in cryptography, they must be resistant to various attacks and non-standard situations and have a sufficiently long sequence period.

Protocols for creating keys in the form of a bit sequence and the mathematical foundations of such protocols are considered in works [21-23], but they only partially consider the creation of permutation matrices, or simple bit sequences.

At the same time, in this paper, we solve the problem of generating a sequence of permutations in their various isomorphic representations, not numbers or bits, as in traditional and well-known HPPs. We can use the latter only as arguments or seeds, based on of which a tuple-flow of vectors is formed, which are essentially permutations, or matrices of permutations, which are an isomorphic representation of permutations.

Since in [15, 17], the issues of coordinating only the main matrix permutation of the general type were considered, and not a sequence (stream) of matrix permutations, the aim of this work is to model and study the processes of creating a sequence of matrix permutations for matrix-algebraic methods of cryptographic transformation in matrix-based systems, as well as to analyze the statistical and correlation characteristics of this sequence.

II. PRESENTATION OF THE MAIN MATERIAL AND RESEARCH RESULTS

Let's consider a situation where we use matrix permutations (MP) for cryptographic transformation of data blocks of size 256x256 bytes, which can be represented as either grayscale images or vectors of length 256 bytes (2048 bits). These matrix permutations and the processes of their creation are described in references [10, 15], which also provide detailed instructions on their generation and use for cryptographic transformations.

Since each data block undergoing multiple rounds of cyclic cryptographic transformations requires its own sequence of matrix keys (MK), there is a need to investigate the process of efficient and reliable generation of such a sequence of MK in the form of MP. Let the number of these MK also be equal to 256.

In Fig.1 the results of modeling the process of generating a sequence of MK for such a case, performed in Mathcad using formulas and MP matrices are shown. If the main MK is a randomly generated matrix permutation KPX (Fig. 1), it is uniquely mapped to a vector permutation V_KPX with 256 components, as well as in the form of an image or a matrix of bytes of size 16x16. An important feature is that all 256 intensity levels in this image are unique.

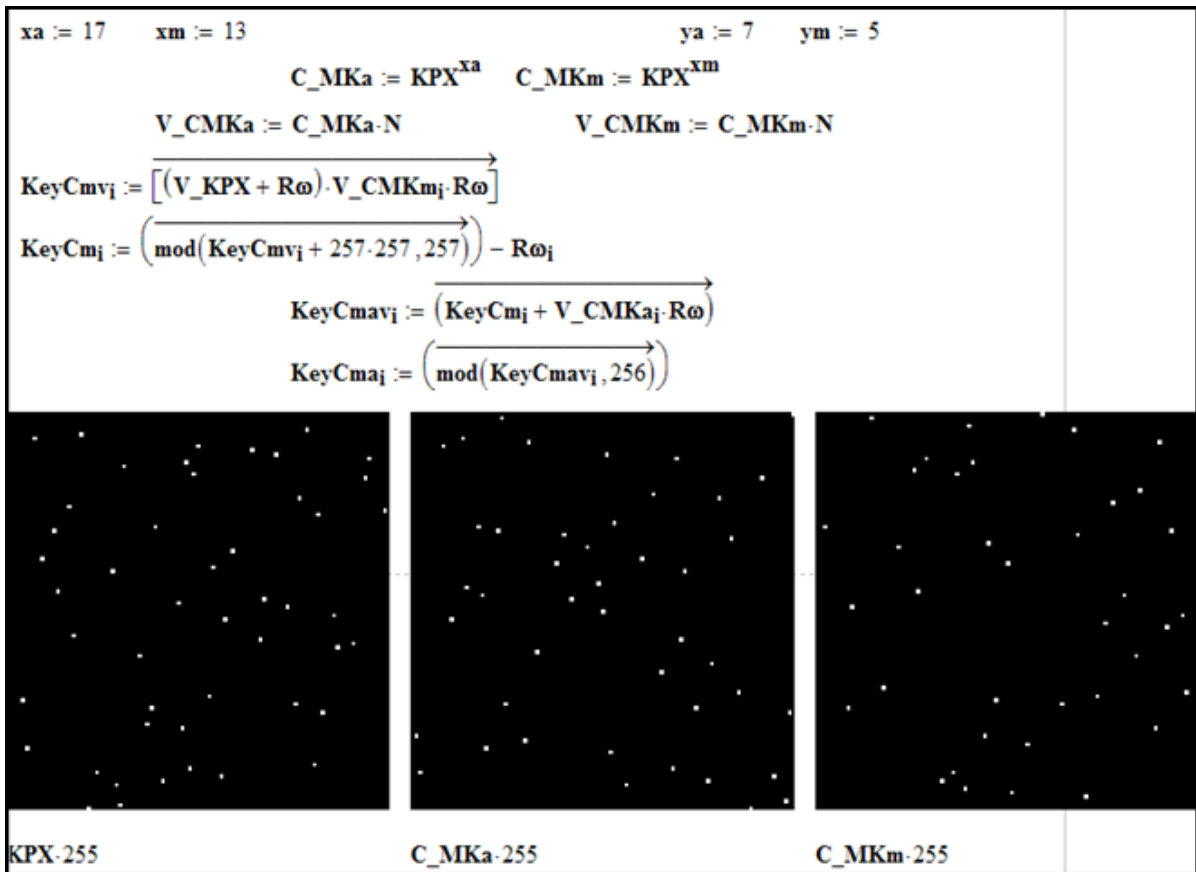


Fig. 1. Results of modeling the processes of generating an set-stream of MKs (MPs).

By utilizing the coordinated scalars x_a and x_m , raised to the power, we obtain two additional matrices, denoted as C_MKa and C_MKm (see Fig. 1), along with corresponding vectors denoted as V_CMKa and V_CMKm . These matrices and vectors, together with the vector V_KPX (which is the vector representation of KPX), are depicted in Fig. 2. Histograms of all these vectors (primary ones!) are horizontal lines, just like the vector representations of the created permutations, which are formed from V_KPX using an affine cipher and a pair of their vector components from vectors V_CMKa and V_CMKm (additional and multiplicative components).

These generated permutations are also represented as binary matrices, for example, $KeyCmaP$ with a size of 256×256 (see Fig. 1), and labeled as $KeyCmaP1-254$.

In Fig. 3 Fragments from Mathcad windows are shown. Since the histograms of all PMs (their vectors) are horizontal lines, and their entropy is equal to 8 bits, crypto-analysis based on them is impossible. In addition, the main and two auxiliary MKs are secret, allowing only parties to the CT to create or have this series of MKs (PMs). In principle, only the master and the aforementioned x_a and x_m scalar keys can be secret or negotiated parties. To study the quality of MKs (PMs) of the created series, to study their properties, we calculated all their possible mutual-correlation and normalized equivalence functions, which in Fig. 4 the Fragments from Mathcad windows are shown.

Note that the obtained results and their comparison also indicate that mutual-equivalence normalized functions are better than mutual-correlation functions. For better perception and more effective transmission of basic MK (PM) and the sequence of created PMs, the latter are converted with the help of software modules into color or black and white image, shown in fig.5 and can go as frames of a video stream (colored image corresponds to three basic MKs).

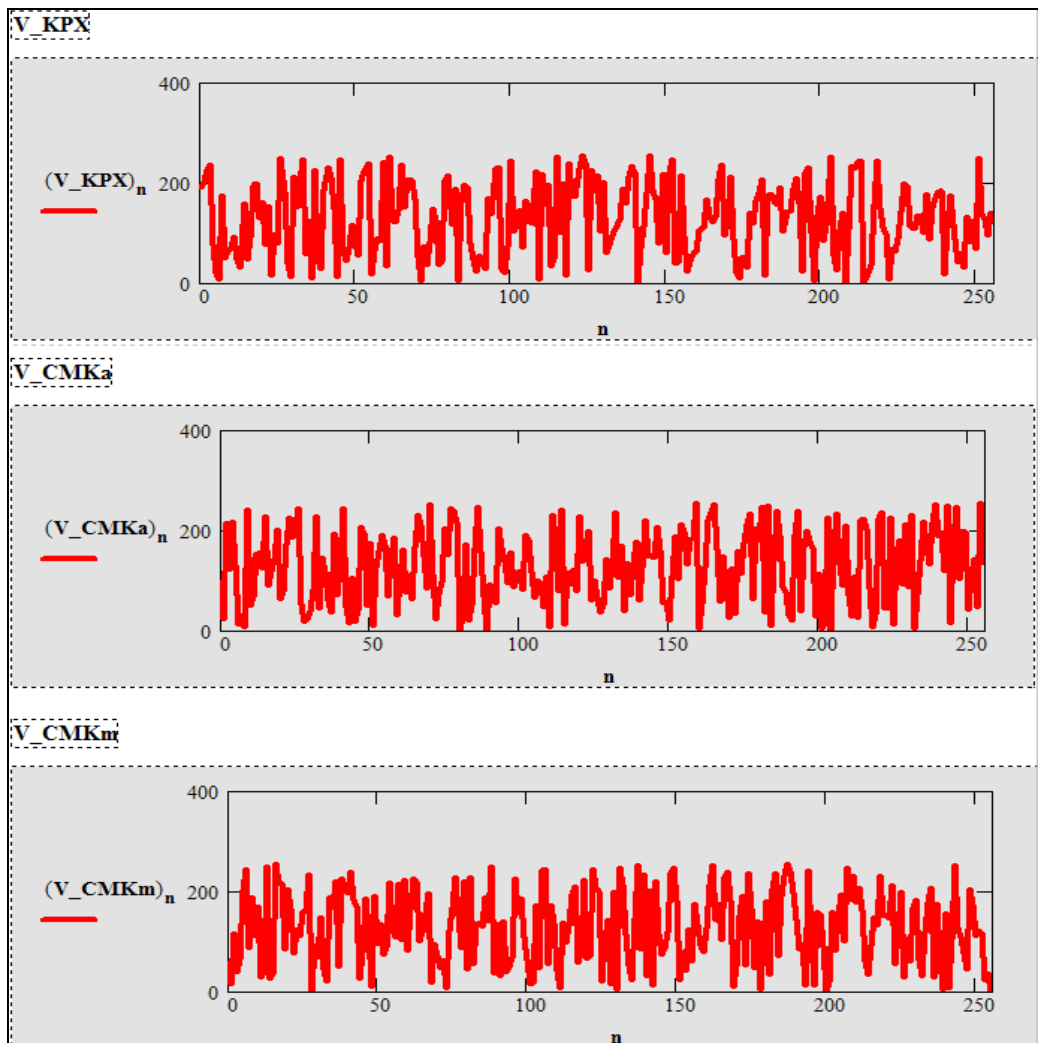


Fig. 2. Basic MKs for forming an array of MKs (MPs) in vector representations.

Since the histograms of all permutation matrix (their vectors) are horizontal lines, and their entropy is 8 bits, this means that cryptanalysis based on them becomes practically impossible. Furthermore, the main key and the two auxiliary keys are strictly confidential, allowing only authorized parties to create or have access to this sequence of matrix keys (MPs). In general, only the main key and the aforementioned scalar keys x_a and x_m can be secret or agreed upon parties.

To assess the quality of the created sequences of matrix keys (MK) or matrix permutations (MP), we calculated all possible normalized cross-correlation and equivalence functions. These results are presented in fragments from the Mathcad program in Fig. 5-6 and confirm the high quality of these sequences. It should be noted that the comparison of the obtained results indicates the superiority of normalized equivalence functions over cross-correlation functions.

Observing Figs 6-7, one can notice that one of the matrix keys (in this experiment, the 200-th) has a similar appearance to another key. However, this can be explained by the fact that in this case, x_m equals "1." This similarity can be easily eliminated by reducing the number of matrix keys in the sequence from 256 to 255, as specified in the chosen modeling and described in this context

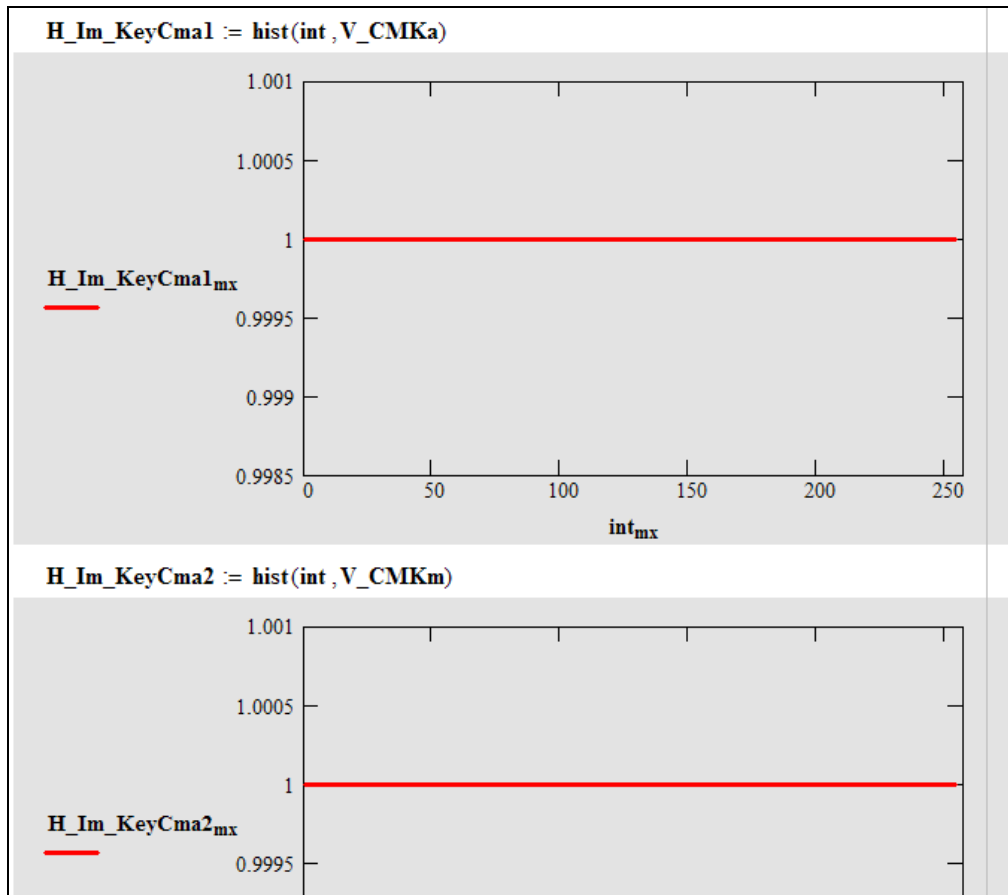


Fig. 3. Histograms of vector representations of basic and some generated MK (MP).

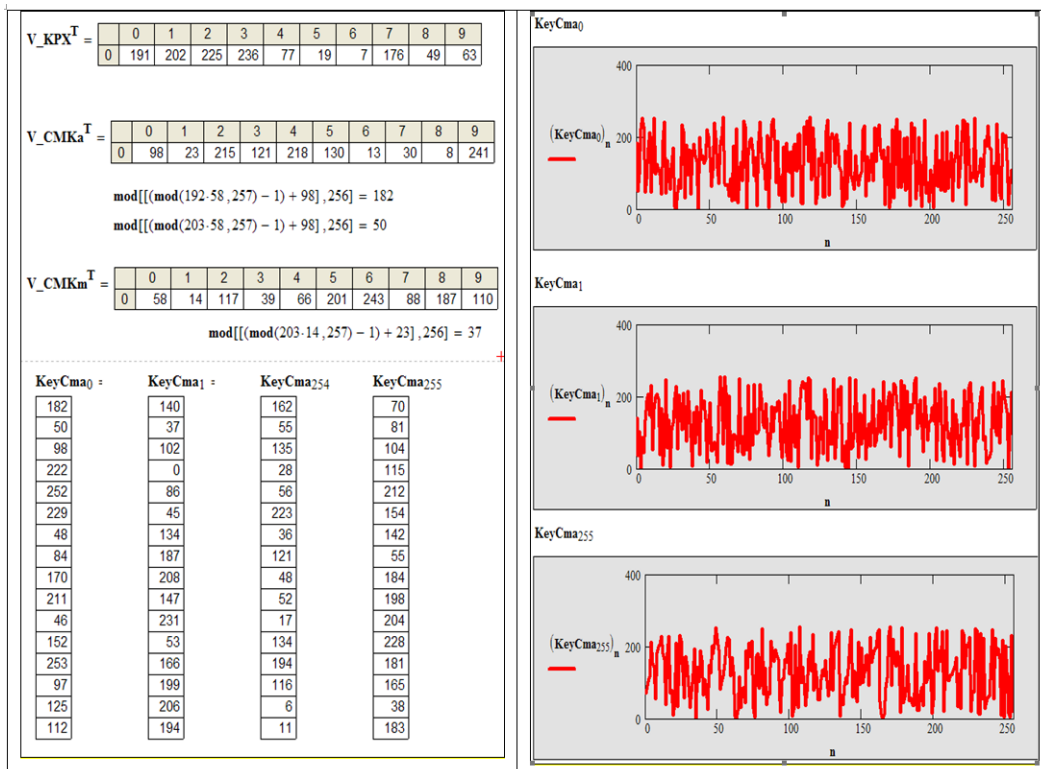


Fig. 4. Mathcad window fragments: one of the key generation procedures (left) and vector representations of some (zero-th, first, 255-th) generated (right) MK (MP).

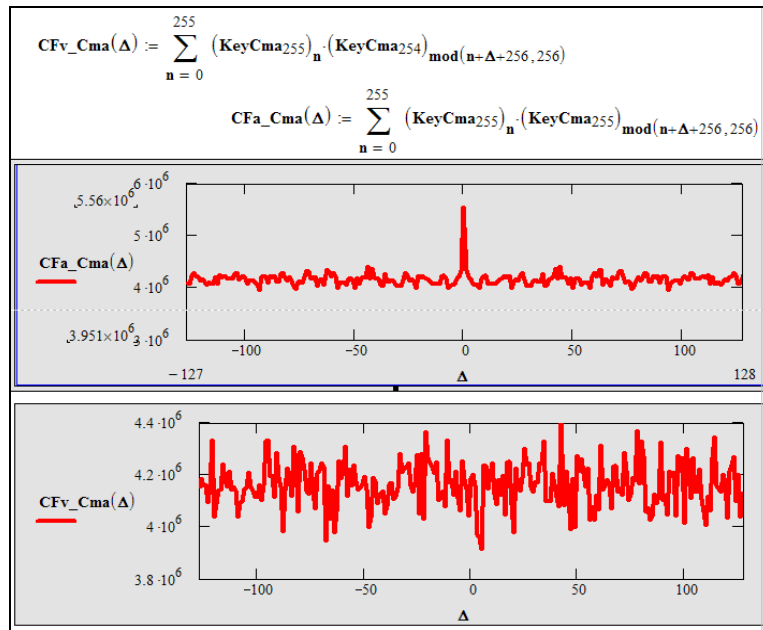


Fig. 5. Formulas from Mathcad window and representation of auto-correlation function CFa_Cma and cross-correlation function CFv_Cma depending on cyclic shift, displacement of elements in the MP vectors.

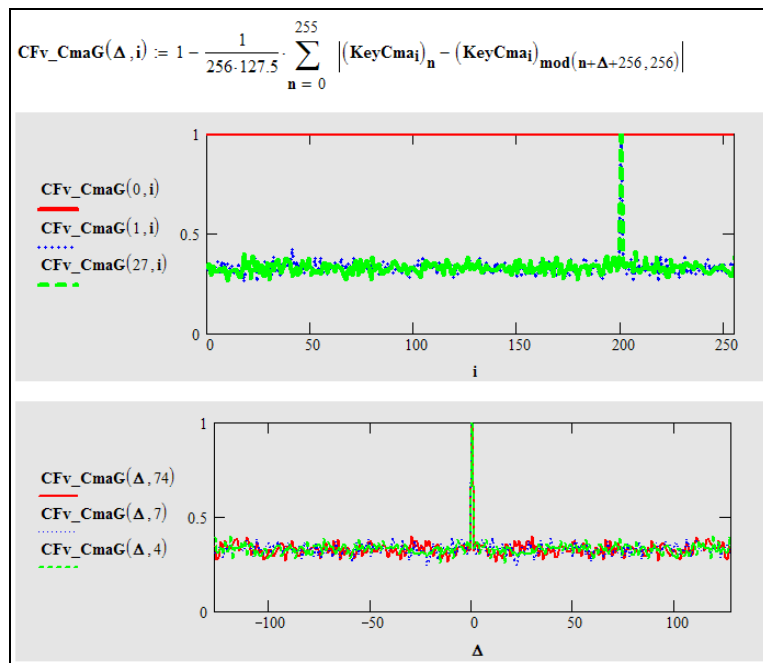


Fig. 6. Formulas from Mathcad window and form of cross-equivalence functions CFv_CmaG depending on the MP number (i) and cyclic shift, displacement of elements in the MP vectors.

For ease of comprehension and more efficient transmission of the main matrix keys (MK) and sequences of created matrix permutations (MP), the latter are transformed into color or grayscale images using software modules, as shown in Fig. 7. These images can be transmitted as frames of a video stream (color image corresponds to the three main matrix keys). The BP_MAPC simulation with the generated keyset was done using Mathcad. Windows for GT simulation will be presented in the report. The essence of BP_MAPC is to apply to blocks, as a set of bytes (PIC_S), procedures for pixel-by-pixel multiplication / addition modulo with MK (direct / inverse). Simulation of the processes of direct / reverse CT TGD, images confirm the correctness of the models.

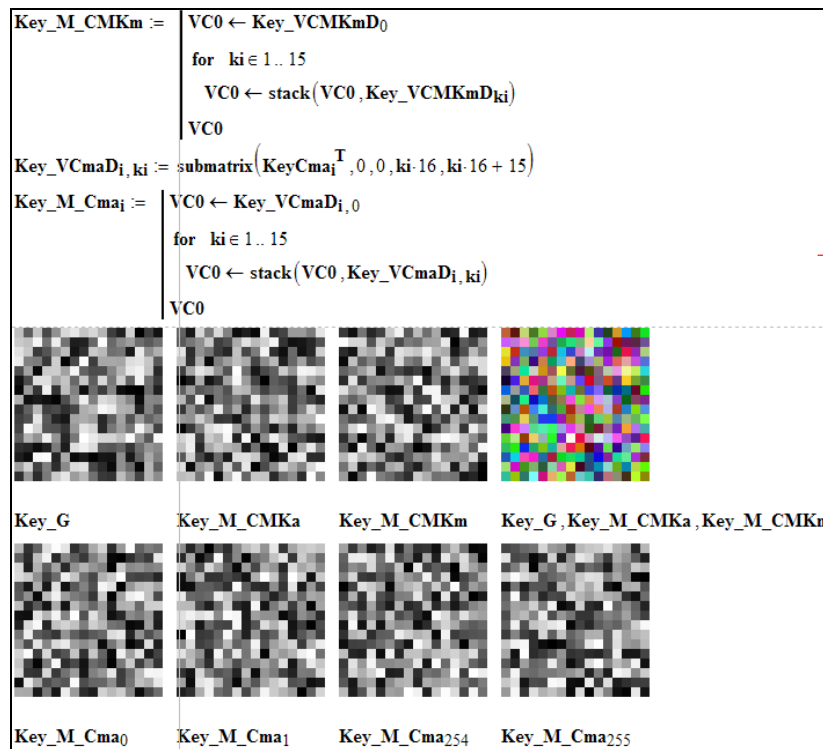


Fig. 7. Mathcad window fragments: Matrix representation of basic MK and a series of MPs.

In order to improve the algorithm, we propose to apply various current MKs to blocks and increase the size of MKs, blocks up to 256x256 bytes. Thus, the idea of BP_MAP ciphers of CT is to use the functional dependencies of their parameters on block indices and additional vector keys (VC). MP in the generally accepted form should be square with elements $N \times N$ ("0" or "1"), where $N = 2^{16}$. The power of the set of MPs, that is, their number, is estimated as $N!$, which gives huge values. Each block address can be represented by two bytes denoting two block coordinates. In [24, 25] questions were considered of creating by the parties a secret MMK of type P with isomorphic representations and the synthesis of set number of sub-keys of a similar type from it (will be covered in the report).

III. DISCUSSION

Within the framework of this work, the set goal has been achieved, namely, a method for generating a stream of a special type of keys for masking video frames is described. At the same time, some aspects important for further expansion of areas of use remained unexplored. This includes testing the created sequence and its resistance to various types of attacks.

IV. FUTURE RESEARCH

In future research, it will be possible to address these important questions and perform a series of simulations to test the quality of the generated streams for the masking processes of real video files.

V. CONCLUSION

A method of generating a series of matrix keys in the form of permutation matrices and their isomorphic representations, which are necessary for multi-page, block, matrix affine-permutation algorithms and matrix-algebraic models of cryptographic transformations, is proposed. The method is modeled in Mathcad. The properties of the pseudo-random sequence of generated matrix keys in the form of permutation matrices are investigated using mutual correlation and equivalence

normalized functions, and the advantages of equivalence versus correlation are shown. The obtained results confirm the adequacy and reliability of this method.

REFERENCES

- [1] Ferguson N. and Schneier B. Practical Cryptography. John Wiley & Sons.2003.432 p..
- [2] Horbenko I.D., Horbenko Yu.I. Prykladna kryptolohiya. Teoriya. Praktyka. Zastosuvannya. Monohrafiya I.D. Horbenko. – Kharkiv: Fort, 2012. – 878 s.
- [3] Yemets' V. Suchasna kryptohrafiya: Osnovni ponyattya / V. Yemets', A. Mel'nyk, R. Popovych. – L'viv: BaK, 2003. – 144 s.: il.
- [4] Gorbenko I. D., Dolgov V.I., Rublinetskii V.I., Korovkin K.V. Methods of Information Protection in Communications Systems and Methods of Their Cryptanalysis //Telecommunications and Radio Engineering. 1998. Volume 52, Issue 4, pp. 89-96.
- [5] Hu, Z., Gnatyuk, S., Okhrimenko, T., Tynymbayev, S., & Iavich, M. (2020). High-Speed and secure PRNG for cryptographic applications. International Journal of Computer Network and Information Security, 12(3), 1–10. <https://doi.org/10.5815/ijcnis.2020.03.01>
- [6] Security Comparison Between Wi-Fi 6 and 5G. <https://forum.huawei.com/enterprise/en/securitycomparison-between-wi-fi-6-and-5g/thread/615836-869>
- [7] Mcginthy, J. M., & Michaels, A. J. (2019). Further analysis of prng-based key derivation functions. IEEE Access, 7, 95978–95986. <https://doi.org/10.1109/access.2019.2928768>
- [8] Gorbenko I., Kuznetsov A., Lutsenko M. and Ivanenko D. The research of modern stream ciphers //4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkov, 2017, pp. 207-210.
- [9] Kuznetsov O., M. Lutsenko and D. Ivanenko, "Strumok stream cipher: Specification and basic properties, "2016 Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T), Kharkiv, 2016, pp. 59-62
- [10] V.G. Krasilenko, V.M. Dubchak, "Cryptographic Transformations of Images based on Matrix Models of Permutations with Matrix-Bit-Map Decomposition and their Modeling," Bulletin of Khm. National University. Technical sciences, 2014, no. 1, pp. 74-79
- [11] V.G. Krasilenko, D.V. Nikitovich, "Modeling and Research of Cryptographic Transformations of Images based on their Matrix-Bit-Map Decomposition and Matrix Models of Permutations with Verification of Integrity," Electronics and Information Technologies, Lviv: National University, 2016, vol. 6, pp. 111-127.
- [12] A.Ya. Bilets'kyi, A.A. Bilets'kyi, D.A. Stetsenko, "Modyfikovanyy matrychnyy asymetrychnyy kryptohrafichnyy alhorytm Diffi-Khellmana," Shtuchnyy intelekt, 2010, № 3, s. 697-705
- [13] V.G. Krasilenko, A.A. Lazarev, D.V. Nikitovich, "The Block Parametric Matrix Affine-Permutation Ciphers (BP_MAPCs) with Isomorphic Representations and their Research," Actual Problems of Information Systems and Technologies, 2020, pp. 270-282.
- [14] V.G. Krasilenko, D.V. Nikitovich, "Modelyuvannya storinkovykh kryptohrafichnykh peretvoren' masyviv kol'orovykh zobrazhen' na osnovi matrychnykh modeley ta perestanyok," «Informatsiyno-komp'yuterni tekhnolohiyi – 2018»: Zbirnyk tez dopovidey IX Mizhnarodnoyi NTK, 20-21 kvitnya 2018 roku, Zhytomyr: Vyd. O. O. Yevenok, 2018, s. 73-77.
- [15] V. G. Krasilenko, A. A. Lazarev, and D. V. Nikitovich, "Matrix Models of Cryptographic Transformations of Video Images Transmitted from Aerial-Mobile Robotic Systems," in Control and Signal Processing Applications for Mobile and Aerial Robotic Systems. Hershey, PA: IGI Global, 2020, pp. 170-214.
- [16] Yu. I. Horbenko ta I. D. Horbenko, "Infrastruktury vidkrytykh klyuchiv. Systemy ETSP. Teoriya ta praktyka. Monohrafiya." Kharkiv: Fort, 2010, 593 p.
- [17] V. G. Krasilenko and D. V. Nikitovich, "Modeling of multi-step and multi-stage secret matrix key matching protocols," in "Computer-integrated technologies: education, science, production": scientific journal. Lutsk: LNTU, 2017, vol. 26, pp. 111-120. [Online]. Available: [<http://ki.lutsk-ntu.com.ua/node/134/section/27>]
- [18] M. A. Dabbah, W. L. Woo, and S. S. Dlay, "Secure Authentication for Face Recognition," presented at Computational Intelligence in Image and Signal Processing, 2007. CIISP 2007. IEEE Symposium on, 2007.
- [19] M. Kutter, F. Jordan, and F. Bossen, "Digital Signature Of Color Images Using Amplitude Modulation," in Proc. Of the SPIE Storage and Retrieval for Image and Video Databases, 1997, vol. 3022, pp. 518-526.
- [20] R. N. Kvyetnyy, Ye. O. Tytarchuk, and A. A. Hurzhiy, "Metod ta alhorytm obminu klyuchamy sered hrup korystuvachiv na osnovi asymetrychnykh shyfriv ECC ta RSA," in "Informatsiyni tekhnolohiyi ta komp'yuterna inzheneriya," 2016, no. 3, pp. 38-43.
- [21] V. Luzhets'kyi and I. Horbenko, "Metody shyfruvannya na osnovi perestanyky blokiv zminnoyi dovzhyny," in "Zakhyst informatsiyi," 2015, vol. 17, no. 2, pp. 169-175.
- [22] Yu. I. Hrytsyuk and P. Yu. Hrytsyuk, "Matematychni osnovy protsesu heneruvannya klyuchiv perestavyannya z vykorystanniam shyfru Kardano," in "Naukovyy visnyk NLTU Ukrayiny," 2015, vol. 25.10, pp. 311-323.
- [23] Yu. I. Hrytsyuk and P. Yu. Hrytsyuk, "Metody i zasoby heneruvannya QP-matryts' Fibonachchi-klyuchiv dlya realizatsiyi kryptohrafichnykh peretvoren'," in "Naukovyy visnyk NLTU Ukrayiny," 2015, vol. 25.6, pp. 334-351.
- [24] V. G. Krasilenko, N. P. Yurchuk, and D. V. Nikitovich, "Zastosuvannya izomorfnykh matrychnykh predstavlen' dlya modelyuvannya protokolu uz-hodzhennya sekretnykh klyuchiv-perestanyok znachnoyi rozmirnosti," in "Visnyk Khmel'nyts'koho natsional'noho universytetu. Tekhnichni nauky," Khmel'nyts'kyi, 2021, vol. № 2, pp. 78-88.
- [25] V. G. Krasilenko and D. V. Nikitovich, "Modeling of methods for generating flows of matrix permutations of significant dimension for cryptographic transformations of images," in Abstracts of the II All-Ukrainian STC Computer Technologies: Innovations, Problems, Solutions. Zhytomyr: Zhytomyr Polytechnic, 2019, pp. 67-77.

Methods and algorithms for forecasting the effective use of energy resources for household consumers

Roman Dyndyn, Ihor Lazarovych, Serhii Ishcheriakov

*Department of Information Technology
Vasyl Stefanyk Precarpathian National University
Ivano-Frankivsk, Ukraine*

Abstract—The scientific work is devoted to the improvement of models for forecasting the use of energy resources based on machine learning. The study included the application of various methods and algorithms of machine learning, among which the approach based on random forests turned out to be the most effective. As a result of the application of this method of prediction, development, and application of additional features, it was possible to achieve an improvement in accuracy compared to other algorithms. This shows the high potential of using machine learning to optimize the use of energy resources in household conditions.

Keywords—*Machine Learning, Random Forest, Forecasting the use of energy resources, Feature engineering.*

I. INTRODUCTION

In the modern world, the problems of energy efficiency and rational use of energy resources are relevant to the development of society. The need for accurate forecasting of utility consumption is due to many factors. From an economic point of view, such forecasting will improve the efficiency of households. It will allow you to plan and manage your expenses. If necessary, optimize the use of housing and communal resources. This is even more useful when there are several households in the property, and then the management of all utility services becomes exponentially more difficult. Taking into account the growth of the number of household consumers and the increase in their needs for electricity, gas, water, and other energy resources, it is necessary to ensure their efficient and rational use.

II. RELATED WORK

In recent years, interest in the use of machine learning methods for the classification and regression of energy consumption has increased. Konstantin Hopf [1] investigates the problem of classification of the efficiency of energy consumption based on data on monthly electricity consumption. He developed new geographic and weather-based features and then trained and compared single, ensemble, and hybrid models with single-family home data. In works [2, 3], various models of deep learning models were used for data processing and forecasting: Deep Fully Connected, Convolutional, and Long Short-Term Memory Neural Networks. In particular, the energy consumption forecasting approach proposed by the authors of [2] can predict monthly energy consumption with an absolute error of 31.83 kWh and a relative error of 17.29%. A review of the literature indicates a good potential for the use of machine tools for forecasting the energy consumption of residential buildings. In particular, neural network methods, support vector machine, and ensemble methods (mainly XGBoost and random forests) were the most popular algorithms due to their powerful capabilities to solve complex nonlinear dependencies between different parameters.

III.METHODOLOGY

Since the data on the consumption of communal resources for households is confidential information, such data are not publicly available. Therefore, to collect such data, a questionnaire was developed that contains the following questions: date, gas and electricity consumption during a given month, household area, number of people living in a given household, settlement. Using this method, gas and electricity consumption data were collected for 2 years for 7 households.

Spatial-temporal measurements of consumption data allow additional data sources to be added to reveal hidden information from environmental data. Environmental information, such as weather observations, has a significant impact on energy use and can help remove environment-related variations in the data (e.g., energy consumption suddenly increases as temperature drops). Because the captured data is time-stamped and geo-located, it can be associated with weather conditions. Previous research [4] found a positive correlation between electricity consumption and weather data.

Each of the categories of communal services has at least annual data on the consumption of the corresponding resource. Even this small piece of information contains valuable details about customers, but it needs to be further processed to extract information. Annual consumption information is usually derived from manual meter readings. Since the number of days in each month is different, the time intervals of measurements may differ. Thus, the first goal of data preparation is to bring this information together. The second purpose of data processing is to remove redundancy from the data. Namely, the fact that consumption data over successive years are highly correlated since household characteristics and consumption behavior do not change much over time. The observation that consumption values are correlated across several years is supported by theory, as household characteristics and people's living situations — the main causes of energy or water consumption — do not change much over time [5]. A third purpose of feature extraction from annual data is to include information about other customers in the variables for analysis. To understand this, it is useful to set the consumption of each household in relation to the consumption of neighboring households. To achieve the three objectives of data preparation, the following characteristics are used:

- normalized consumption per day;
- annual consumption trend;
- the deviation of the mean (logarithm) consumption to the mean consumption of all neighbors, expressed as a multiple of the standard deviations, to quantify the household's consumption deviation from its neighborhood.

Investigating the seasonality of energy consumption, one can see a clear annual cycle. When dealing with a characteristic such as the "Month", which has an annual seasonality, it is customary to present it in such a way as to reflect its cyclical nature. One effective way to do this is to use sine and cosine transformations:

$$MonthSin = \sin\left(\frac{2*\pi*m}{12}\right), \quad (1)$$

$$MonthCos = \cos\left(\frac{2*\pi*m}{12}\right), \quad (2)$$

where m - month index.

Incorporating these transformed features into our machine learning model should help capture annual seasonality more effectively than using raw "Month" values. This approach prevents the model from treating December and January as very distant points, although they are close in a cyclical sense.

Figure 1 shows all the features used to train machine learning models.

Area	Person	Lag1	Lag2	Correlation	MeanDailyConsumption	DailyConsumptionDeviation	Trend	Month_Sin	Month_Cos	Year_Sin	Year_Cos
44.0	2	109	107	-0.511820	3.817982	0.101135	-0.155979	5.000000e-01	-8.660254e-01	-6.447061e-13	1.0
44.0	2	107	121	-0.511820	3.817982	0.101135	-0.155979	8.660254e-01	-5.000000e-01	-6.447061e-13	1.0
44.0	2	121	121	-0.511820	3.817982	0.101135	-0.155979	1.000000e+00	6.123234e-17	-6.447061e-13	1.0
44.0	2	121	123	-0.511820	3.817982	0.101135	-0.155979	8.660254e-01	5.000000e-01	-6.447061e-13	1.0
44.0	2	123	104	-0.511820	3.817982	0.101135	-0.155979	5.000000e-01	8.660254e-01	-6.447061e-13	1.0
...
60.9	2	107	116	-0.564552	4.336590	0.236746	-0.089648	-1.000000e+00	-1.836970e-16	-1.305266e-12	1.0
60.9	2	116	132	-0.564552	4.336590	0.236746	-0.089648	8.660254e-01	-5.000000e-01	-1.305266e-12	1.0
60.9	2	132	138	-0.564552	4.336590	0.236746	-0.089648	-5.000000e-01	-8.660254e-01	-1.305266e-12	1.0
60.9	2	138	138	-0.564552	4.336590	0.236746	-0.089648	1.224647e-16	-1.000000e+00	-1.468363e-13	1.0
60.9	2	138	138	-0.564552	4.336590	0.236746	-0.089648	5.000000e-01	-8.660254e-01	-1.468363e-13	1.0

Figure 1 – List of all the features used for forecasting

Next, a study was conducted on how different machine learning methods cope with the forecasting of electricity and gas consumption. A cross-validation method was used to calculate the performance indicators, namely 80% of the data was used for training and 20% for testing. Indicators were calculated using mean absolute error (MAE) and mean percentage error (MAPE):

$$MAE = \frac{\sum_{i=1}^n |actual_i - forecast_i|}{n}, \quad (3)$$

$$MAPE = \frac{\sum_{i=1}^n \frac{|actual_i - forecast_i|}{|actual_i|}}{n} * 100, \quad (4)$$

where n – number of values, $actual_i$ – actual value, $forecast_i$ – forecast value.

IV. RESULTS

Table 1 shows the cross-validation results of various machine learning models.

Table 1 – Cross-validation results of electricity usage forecasting

Method	MAE	MAPE
Linear regression	18.154	18.43%
Decision tree	22.333	22.25%
Random forest	17.137	17.13%
XGBoost	20.763	20.43%
Support vector machine	19.651	24.72%
Feedforward neural network	22.734	23.19%

As we can see, random forests showed the best result on the prepared data. Figure 2 shows the results of 1 month-ahead forecasting for one consumer for the period from November 2022 to October 2023.

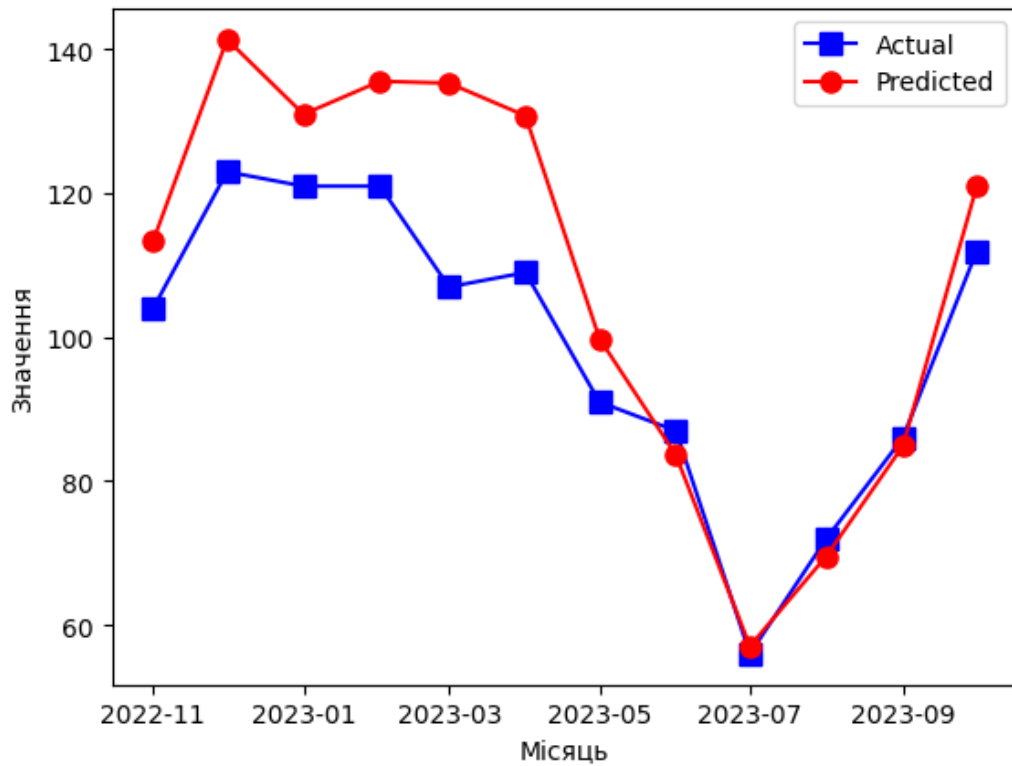


Figure 2– Graph of predicted and actual electricity consumption for one user.

Forecasting gas usage showed a slightly worse result. The random forest method was also the most accurate with a result of 24.2%. In fig. 3 shows the forecast of gas consumption for one consumer for the period from November 2022 to October 2023.

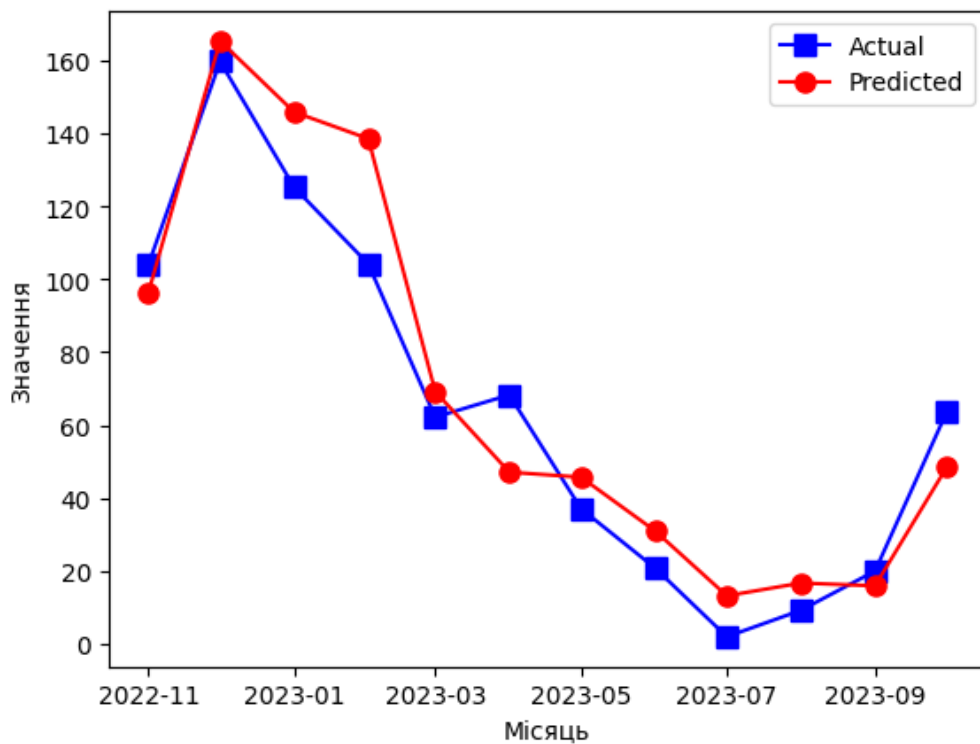


Figure 3– Graph of predicted and actual gas consumption for one user.

V. DISCUSSION

A study of the accuracy of forecasting for the given models was conducted, which allowed us to conclude that in the conditions of a limited amount of dataset for model training, the minimum forecasting error is given by the method of random forests. Due to feature engineering [6, 7], managed to improve the accuracy of electricity usage prediction by 1% compared to [2], using the same dataset for both models. It is also worth noting that fewer computing resources are used to train the model using random forests and make predictions.

VI. FUTURE RESEARCH

Considering the small amount of data sample for model training, it is necessary to investigate the behavior of the proposed approach on more voluminous data. It is also advisable to investigate in more detail the influence of various holiday periods, such as Easter, on the consumption of energy resources.

VII. CONCLUSION

The article investigates the accuracy of energy resource consumption forecasting using various machine learning models under the conditions of a limited dataset. Using feature engineering, accuracy was improved, and the use of random forests reduced the computational complexity of the prediction.

REFERENCES

- [1] K. Hopf, "Data sources in organizations and extraction of predictor variables," in *Predictive Analytics for Energy Efficiency and Energy Retailing*, University of Bamberg Press, 2019, pp. 39-82, doi: 10.20378/irbo-54833.
- [2] R. F. Berriel, A. T. Lopes, A. Rodrigues, F. M. Varejão and T. Oliveira-Santos, "Monthly energy consumption forecast: A deep learning approach," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017, pp. 4283-4290, doi: 10.1109/IJCNN.2017.7966398.
- [3] M. Kozlenko, I. Lazarovych, and M. Kuz, "Deep learning approach to signal processing in infocommunications," in *Proc. 4th International Scientific and Practical Conference on Applied Systems and Technologies in the Information Society (AISTIS)*, V. Pleskach and V. Mironova, Eds. Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, Sept. 30, 2020, pp. 81-82, doi: 10.5281/zenodo.4482757.
- [4] L. Hernández *et al.*, "A study of the relationship between weather variables and electric power demand inside a smart grid/smart world framework," *Sensors*, vol. 12, no. 9, pp. 11571–11591, 2012. doi: 10.3390/s120911571.
- [5] Wells, William D. and George Gubar (1966). "Life Cycle Concept in Marketing Research." In *Journal of Marketing Research* 3.4, pp. 355–363. doi: 10.2307/3149851.
- [6] M. Kuhn, K. Johnson, *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton ; London ; New York: CRC Press, Taylor et Francis Group, 2021.
- [7] O. Solovei, "Selection of data preprocessing technique for imbalanced dataset," in *2022 International Conference on Innovative Solutions in Software Engineering (ICISSE)*, Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, Nov. 29-30, 2022, pp. 48-50, doi: 10.5281/zenodo.7502536.

Вибір нефункціональних вимог при розробці програмного забезпечення

Володимир Кімак

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація—В даному дослідженні проаналізовано процес вибору нефункціональних вимог при розробці програмного забезпечення. Запропоновано новий метод для визначення джерела нефункціональних вимог, який базується на відомих параметрах проекту.

Ключові слова — *Розробка програмного забезпечення, Збір та аналіз вимог, Нефункціональні вимоги, Параметри та обмеження проекту, Якість програмного забезпечення.*

I. ВСТУП

Вимоги до програмного забезпечення (англ. Software requirements) зосереджуються на потребах, які повинні вирішуватися цим програмним забезпеченням його потенційними користувачами, а також містить властивості, якості та конкретні функції [1]. При розробці програмного забезпечення найпершим кроком є збір та визначення вимог з найрізноманітніших джерел інформації, таких як матеріали бізнес-аналітиків, декомпозиція завдань та функцій, технічні співбесіди з бізнес-експертами, або відповідальними особами за певний набір завдань [2]. Це один з найважливіших кроків, оскільки весь продукт залежить від зібраних вимог. Після того, як вимоги будуть зібрані, вони аналізуються і документуються в специфікації вимог до програмного забезпечення. Вимоги можна поділити на дві групи, такі як функціональні та нефункціональні вимоги.

Функціональні вимоги (англ. Functional requirements) - це перелік функцій або сервісів, які повинна надавати система, а також обмежень на дані і поведінку системи при їхньому виконанні. Специфікація функціональних вимог - опис функцій та їхніх властивостей, які не містять у собі протиріч і виключень. Вони відповідають на питання “що” повинна робити система.

Нефункціональні вимоги (англ. Non-functional requirements) визначають якісні характеристики системи (наприклад, безпека, швидкодія, зручність користування) та обмеження накладені на неї (наприклад, середовище, сертифікації). Ці вимоги характеризують принципи взаємодії всередині системи та інтеграції з іншими системами, а також визначають показники часу роботи, захисту даних і досягнення якості з урахуванням рекомендацій використовуваного стандарту. Вони виражаються як у абсолютних значеннях (наприклад, час очікування відповіді, кількість клієнтів, що обслуговуються і ін.) так і відсоткових (наприклад, показник надійності чи гарантований показник роботоздатності (англ. Service level agreement) [3]). Вони відповідають на питання “як” повинна працювати система.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Є декілька різних класифікацій нефункціональних вимог. Наведу одну з найпопулярніших класифікацій [4]. В ній визначені наступні критерії вимог:

- продуктивність (performance) –
- масштабованість (scalability)

- портативність (portability)
- сумісність (compatibility)
- надійність (reliability)
- підтримуваність (maintainability)
- доступність (availability)
- безпека (security)
- локалізація (localization)
- зручність у користуванні (usability)

Для інших класифікацій можуть використовуватись як критерії типи/категорії вимог згідно [5] чи [6].

Потрібно зауважити що не всі критерії однаково важливі для кожного проекту. У кожному конкретному випадку різні критерії отримують різний пріоритет у залежності від специфіки цього проекту чи випадку. Це так звані компроміси (англ. trade-off), які приймаються при розробці архітектури проекту. Один зі способів визначення важливості критеріїв та оцінки їхньої ваги описаний у [1]. Ідея полягає в тому, щоб відобразити критерії у вигляді матриці, яка позначає вимоги, які мають пріоритет над іншими, коли вони мають суперечливий вплив на систему. Наприклад, продуктивність і безпека системи, завжди між собою конфліктують, так як перевірка та валідація витрачають додатковий час, і потрібно вирішити у кожному випадку, де потрібно вибрати продуктивність, а де безпеку.

На даний момент розроблено багато підходів щоб визначити ті критерії, які будуть реалізовані у конкретному проекті. Наведу декілька варіантів, які зустрічаються на реальних проектах.

- Підхід у вигляді опитувальника і майстер класу зображеного на рис. 1 з визначення атрибутів якості [7]. Цей підхід допомагає відповісти на багато питань, але він є досить затратним по часу і ресурсам.

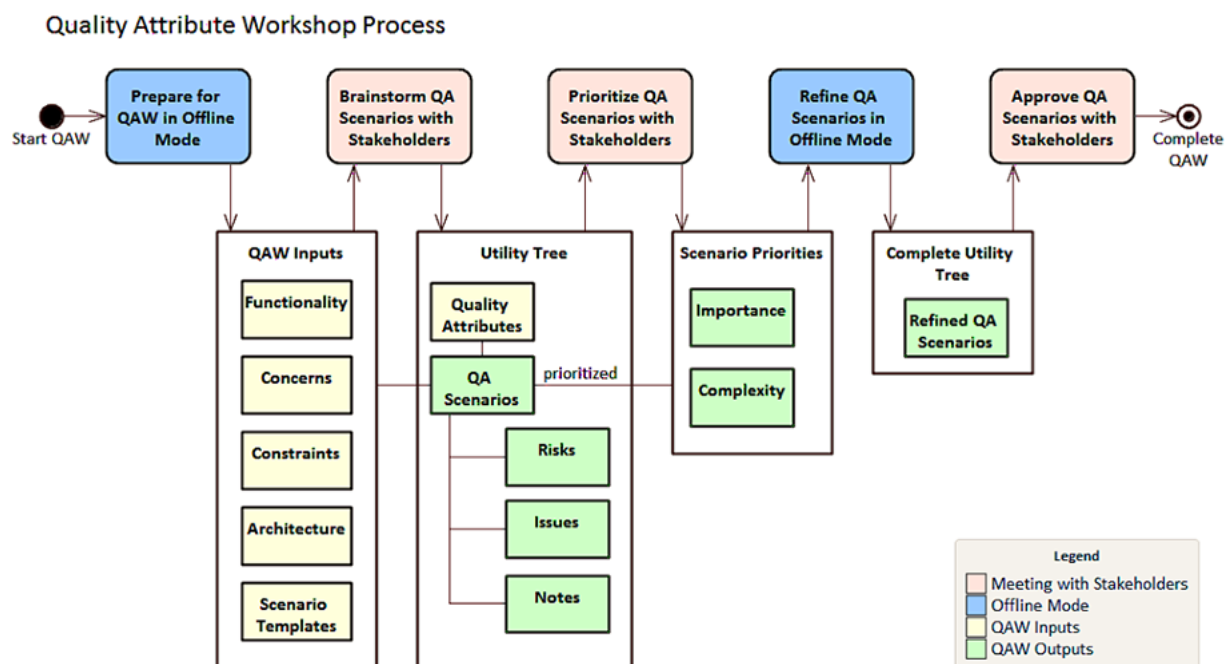


Рис. 1 – Діаграма майстер-класу з вибору та пріоритезації нефункціональних вимог [7]

- Підхід, коли нефункціональні вимоги формуються під час реалізації проекту, але нажаль цей підхід в результаті дає тільки частковий результат і досить часто проект переписується не дійшовши до фази впровадження.
- Підхід, при якому використовуються стандарти [5] чи [6] для вибору вимог за допомогою опитування всіх зацікавлених сторін.

- Підхід, при якому нефункціональні вимоги описуються як функціональні у вигляді історій користувачів, наприклад «Я, як користувач, хочу щоб сайт завантажувався швидко, що мені дозволить зекономити час для отримання інформації», проблема даного підходу в тому що потрібно додатково визначити як вимірювати коректність та правильний результат цих історій.

III. МЕТОДИ

В цій роботі пропонується алгоритм для визначення джерела нефункціональних вимог, який базується на відомих параметрах проекту. Як відомо, кожен проект є відносно унікальним з точки зору розробки, оскільки він містить певний визначений набір параметрів, серед яких варто виділити наступні:

- специфіка домену та контексту застосування системи;
- потреби та очікування клієнтів;
- бюджет, терміни та ресурси, виділені для розробки;
- ризики, що пов'язані з недотриманням або невиконанням нефункціональних вимог.

На основі цих параметрів ми можемо визначити найбільш придатне джерело вимог. Наприклад, при малому бюджеті та коротких термінах неможливо провести повноцінний майстер-клас чи використати готовий стандарт, відповідно доведеться впроваджувати нефункціональні вимоги на стадії розробки проекту.

Іншим прикладом може бути робота з доменом, що вимагає строгих функціональних вимог, зразком такого домену є банківський, державний та інші. В такому випадку необхідно проводити повноцінний майстер-клас.

На рисунку 2 зображена схема алгоритму, що розроблений в цій роботі. Він починається з визначення параметрів проекту, для якого виконується планування. Параметрами можуть виступати доменна область, очікування клієнтів, бюджет, ризики та інші. Для кожного з цих параметрів є визначена вага, яка множиться на значення параметру. Вага потрібна тому, що параметри не є взаємозамінними та рівноцінними. Наприклад, бюджет і терміни мають велику вагу, оскільки при виході за межі бюджету та термінів існує дуже великий ризик не випустити продукт в принципі.

Після застосування ваг потрібно просумувати всі отримані значення та визначити коефіцієнт. Чим менший коефіцієнт тим простішим буде джерело нефункціональних метрик. У даному випадку простота відслідковується наступним чином:

- визначення нефункціональних вимог під час розробки – найпростіше джерело;
- консультація з зацікавленими сторонами та використання готових стандартів;
- проведення майстер-класу – найскладніше джерело.

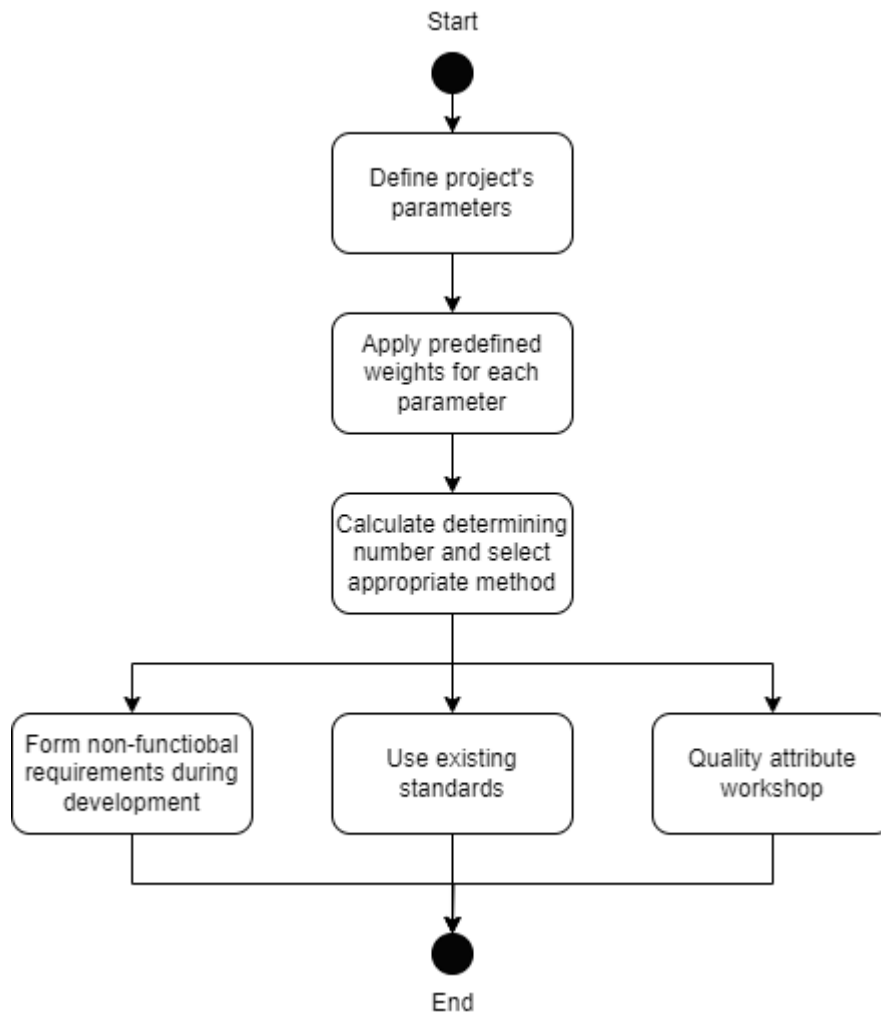


Рис. 2 – метод вибору джерела нефункціональних вимог в залежності від параметрів проекту

Розроблений метод дозволяє ефективно визначати найбільш підходяще джерело нефункціональних вимог для конкретного проекту на стадії планування.

IV. РЕЗУЛЬТАТИ

В даній роботі було проаналізовано існуючі підходи для розробки нефункціональних вимог. Описано новий підхід та запропоновано метод для визначення джерела нефункціональних вимог, який базується на відомих параметрах проекту.

V. ОБГОВОРЕННЯ

Використання запропонованого методу для визначення джерела дозволяє досить легко зрозуміти, як варто вибирати нефункціональні вимоги для проекту, так як він заснований на тих параметрах проекту, які відомі практично зразу, такі як домен проекту та його бюджет і терміни, або описані у проектній документації, такі як функціональні вимоги. Це означає, що впровадження цього підходу на практиці не займе багато часу і дозволить покращити якість розроблюваного програмного забезпечення.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

При подальших дослідженнях можливо покращити підхід шляхом введення нових кроків в алгоритм для покращення його точності, а також можлива подальша розробка таблиці з вагами.

VII. ВИСНОВКИ

Розробка нефункціональних вимог є обов'язковою частиною розробки будь-якого проекту. Вибір підходу для визначення нефункціональних вимог є дуже важливим, оскільки неправильно підібрані нефункціональні вимоги можуть привести до провалу проекту після релізу. Використання найбільш підходящого підходу для визначення нефункціональних вимог дозволяє гарантувати якість проекту після випуску та якісно планувати час для розробки.

ЛІТЕРАТУРА

- [1] K. Wiegers, J.Beatty "Software Requirements", 3rd ed. Washington: Microsoft Press, 2013, pp. 143-162, 263-294
- [2] Yu. Hrytsiuk, "Software Requirements Analysis", Lviv (in Ukrainian), 2018, pp. 16-25
- [3] "What is a service level agreement (SLA)?" ServiceNow
<https://www.servicenow.com/products/itsm/what-is-sla.html> (accessed Nov. 1, 2023)
- [4] "Non-functional Requirements: Examples, Types, How to Approach" Altexsoft.com
<https://www.altexsoft.com/blog/non-functional-requirements/> (accessed Nov. 5, 2023)
- [5] "ISO/IEC 25010:2011" ISO.org
<https://www.iso.org/standard/35733.html> (accessed Nov. 3, 2023)
- [6] "ISO/IEC TS 25011:2017" ISO.org
<https://www.iso.org/standard/35735.html> (accessed Nov. 3, 2023)
- [7] "Quality Attribute Workshop Collection" Cmu.edu
<https://insights.sei.cmu.edu/library/quality-attribute-workshop-collection/> (accessed Nov. 3, 2023)

Моделювання ліній розділу фільтраційних потоків за умови відсутності перетоку із джерелом збурення

Сергій Каштан

*Циклова комісія математики, інформатики та фізики
Відокремлений структурний підрозділ «Рівненський технічний фаховий коледж
Національного університету водного господарства та природокористування»
Рівне, Україна*

Анотація—У роботі досліджуються ідеальні та квазіідеальні поля при наявності джерела поперечних збурень в областях обмежених лініями течії та екіпотенціальними лініями. В залежності від значень потенціалів на екіпотенціальних лініях, можливі різні випадки формування ліній розділу фільтраційних потоків у фізичній області, а отже – побудови області квазікомплексного потенціалу. Наближений розв’язок крайової задачі із ділянкою збурення на лінії течії отримано за допомогою розробленого програмного продукту, побудованого на основі алгоритму, який ґрунтується на почерговій параметризації величин конформних інваріантів, граничних і внутрішніх вузлів сіткової області з використанням ідей методу блочної ітерації.

Ключові слова—*Моделювання, Фільтраційні потоки, Конформні відображення, Обернені задачі, Джерело збурень, Числові методи.*

I. ВСТУП

В багатьох випадках практики формули на знаходження характеристик середовища потребують задання функції швидкості. В реальних фізичних задачах, зокрема в задачах фільтрації, гідрології, часто це зробити або неможливо, або дуже складно. Тому запропоновано підхід, при якому по відомих значеннях потенціалів на кожній із ділянок витоку та втоку рідини в досліджувану область будується відповідна область квазікомплексного потенціалу.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У роботі [1] описано розроблений метод обернених крайових задач (конформних і квазіконформних відображень) для побудови динамічної сітки потенційних та квазіпотенційних полів, побудови різного роду профілів та поля швидкості із паралельним розрахунком різних характеристик процесу. Тут, під оберненням такого роду задач мається на увазі як перехід від прямих задач до задач на конформне відображення відповідної області комплексного потенціалу на вихідну область, так і той факт, що вони містять невідомі параметри при додаткових відомостях про їх розв’язки.

Методику наближення розв’язків крайових еліптичних задач, можна перенести і на крайові задачі, які зводяться до конформних відображень довільної криволінійної області, обмеженої лініями течії та екіпотенціальними лініями на область комплексного потенціалу, що має вигляд многокутника, границя якого складається із вертикальних та горизонтальних відрізків, променів, із декількома невідомими при відповідності вершин.

У роботах [2, 3] розв’язана проблема нелінійного обернення крайових задач на конформні відображення в областях, обмежених лініями течії та екіпотенціальними лініями з потенціалом керування на ділянці однієї із ліній течії. Окрім шуканого потенціалу і функції течії, у результаті побудовано ще й характеристичну функцію, описані усі можливі

характерні випадки формування течії в залежності від заданих значень потенціалу керування, наводяться схеми побудови динамічної сітки, формування відповідних перетоків. При цьому, виділено типи задач (ключових) на знаходження тих значень потенціалу керування, що забезпечують оптимізацію певних функціоналів (витрат, витоків, втоків, перетоків, ін.).

III. ПОСТАНОВКА ЗАДАЧІ

Розглянемо модельну задачу на відшукування гармонічної функції $\varphi = \varphi(x, y)$ (потенціалу) в скінченій однозв'язній криволінійній області (пласт, що піддається певному впливу, напр., зміні, деформації) $G_z = ABMNKCD$ ($z = x + iy$), обмеженій чотирма гладкими кривими (див. рис. 1):

$$AB = \left\{ z : f_1(x, y) = x - 1 - \sin \frac{y}{2} = 0 \right\}, \quad BMNKC = \left\{ z : f_2(x, y) = y - 18 - e^{\sin 0.5x} = 0 \right\},$$

$$CD = \left\{ z : f_3(x, y) = x - 20 - \cos \frac{y}{2} = 0 \right\}, \quad DA = \left\{ z : f_4(x, y) = y - 1 - \cos \frac{x}{3} = 0 \right\}$$

при умовах

$$\varphi|_{AB} = \varphi_* = 0, \quad \varphi|_{CD} = \varphi^* = 1, \quad \varphi|_{MN} = \varphi^\circ,$$

$$\frac{d\varphi}{dn}\Big|_{DA} = \frac{d\varphi}{dn}\Big|_{NKC} = \frac{d\varphi}{dn}\Big|_{BM} = 0,$$

де n – зовнішня нормаль до відповідної кривої [1, 2, 3].

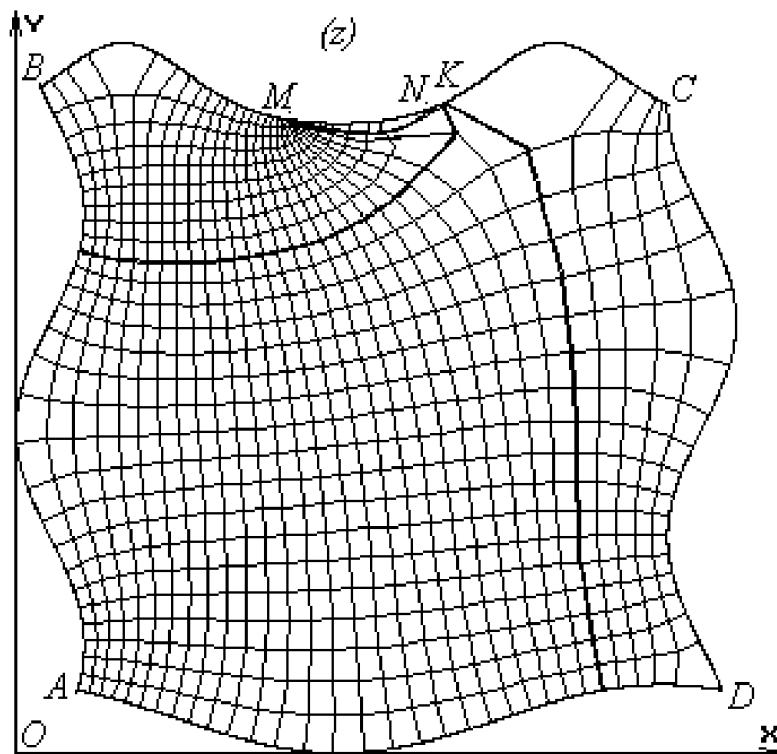


Рис.1. Лінії розділу фільтраційних потоків та сітка фізичної області G_z

Тут значення потенціалу φ° ($\varphi_* < \varphi^\circ < \varphi^*$) є настільки великим, що відсутні перетоки між ділянками MN та CD , а точка K – точка розділу фільтраційних потоків та значення потенціалу у ній $\varphi|_K = \tilde{\varphi}$ – визначаються у процесі розв'язання задачі. У роботі [4] розв'язана

задача, коли точка K співпадає з точкою N та знайдено значення потенціалу $\varphi|_{MN} = \varphi^\circ \approx 0,716$.

IV. МЕТОДИ

Ввівши гармонічну функцію $\psi = \psi(x, y)$ (функцію течії), комплексно спряжену до функції $\varphi = \varphi(x, y)$, і замінивши останні три граничні умови на умови

$$\psi|_{DA} = 0, \quad \psi|_{NKC} = Q_1, \quad \psi|_{BM} = Q_2,$$

де Q_1, Q_2 – величини фільтраційних потоків (невідомі параметри), цю задачу замінимо [1, 2, 3] більш загальною задачею на конформне відображення $\omega = \omega(z) = \varphi(x, y) + i\psi(x, y)$ фізичної області G_z на область комплексного потенціалу $G_\omega = \bigcup_{l=1}^4 G_\omega^{(l)}$,

$$G_\omega^{(1)} = \{(\varphi, \psi) : \tilde{\varphi} \leq \varphi < \varphi^*, 0 < \psi < Q_1\},$$

$$G_\omega^{(2)} = \{(\varphi, \psi) : \varphi_* < \varphi < \tilde{\varphi}, 0 < \psi < Q_1\},$$

$$G_\omega^{(3)} = \{(\varphi, \psi) : \varphi_* < \varphi < \tilde{\varphi}, Q_1 \leq \psi < Q_2\},$$

$$G_\omega^{(4)} = \{(\varphi, \psi) : \tilde{\varphi} < \varphi < \varphi^\circ, Q_1 \leq \psi < Q_2\}.$$

Тоді, відповідна обернена крайова задача на конформне відображення $z = z(\omega) = x(\varphi, \psi) + iy(\varphi, \psi)$ області G_ω на G_z при невідомих $\tilde{\varphi}, Q_1, Q_2$ у диференціальній формі запишеться у вигляді:

$$\begin{cases} \frac{\partial y}{\partial \psi} = \frac{\partial x}{\partial \varphi}, & \frac{\partial x}{\partial \psi} = -\frac{\partial y}{\partial \varphi}, \end{cases} \quad (\varphi, \psi) \in G_\omega;$$

$$\begin{cases} f_1(x(\varphi_*, \psi), y(\varphi_*, \psi)) = 0, & 0 \leq \psi \leq Q_2, \\ f_2(x(\varphi, Q_2), y(\varphi, Q_2)) = 0, & \varphi_* \leq \varphi \leq \varphi^0, \\ f_2(x(\varphi^0, \psi), y(\varphi^0, \psi)) = 0, & Q_1 \leq \psi \leq Q_2, \\ f_2(x(\varphi, Q_1 - 0), y(\varphi, Q_1 - 0)) = 0, & \tilde{\varphi} \leq \varphi \leq \varphi^*, \\ f_2(x(\varphi, Q_1 + 0), y(\varphi, Q_1 + 0)) = 0, & \tilde{\varphi} \leq \varphi \leq \varphi^0, \\ f_3(x(\varphi^*, \psi), y(\varphi^*, \psi)) = 0, & 0 \leq \psi \leq Q_1, \\ f_4(x(\varphi, 0), y(\varphi, 0)) = 0, & \varphi_* \leq \varphi \leq \varphi^*. \end{cases}$$

Наближений розв'язок поставленої задачі отримано за допомогою розробленого програмного продукту, побудованого на основі алгоритму [1], який ґрунтується на почерговій параметризації величин конформних інваріантів, граничних і внутрішніх вузлів сіткової області G_z з використанням ідей методу блочної ітерації [5].

V. РЕЗУЛЬТАТИ

Провівши числові розрахунки на побудову динамічної сітки з точністю $\varepsilon = 10^{-4}$ для параметрів розбиття області $m_0 = 26, m_1 = 27, m = 30, n_0 = 18, n = 26$ (див. рис. 2) при

потенціалі керування $\varphi^\circ = 0,9$, також знайдено значення потенціалу у точці розділу фільтраційних потоків $\tilde{\varphi} \approx 0,86$, витрату вхідного потоку $Q_2 \approx 1,277$, витрату вихідного потоку $Q_1 \approx 0,804$, що дозволяє обчислити витрату джерела збурення $Q_2 - Q_1 \approx 0,473$. При цьому, величина нев'язки квазіконформності отриманої динамічної сітки становить $\varepsilon_* = 10^{-2}$.

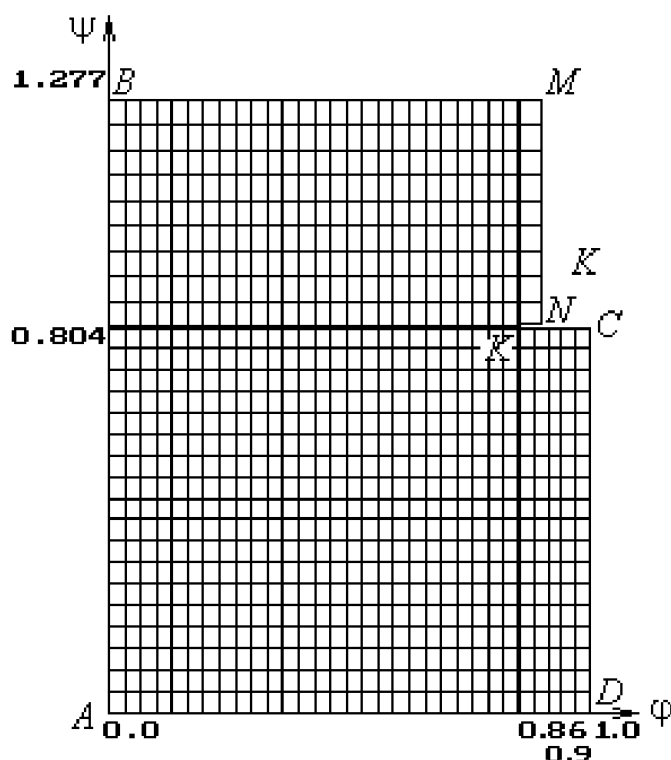


Рис.2. Динамічна сітка області комплексного потенціалу G_ω

VI. ОБГОВОРЕННЯ

Аналізуючи отримані у цій роботі та у [4] результати можна зробити висновок, що лінії розділу фільтраційних потоків за умови відсутності перетоку із джерелом збурення MN будуть утворюватися за подібним профілем у фізичній області G_z . При чому, основна точка їх розділу – точка K – буде пробігати від точки N до точки C , а значення потенціалу $\varphi|_K = \tilde{\varphi}$ знаходитиметься у межах $\varphi|_{MN} \leq \varphi|_K \leq \varphi|_{CD}$, що визначатиме ключові випадки. У цих випадках значення потенціалу φ° знаходиться в процесі розв'язання відповідної задачі при додатковій (“компенсаційній”) умові про відоме положення точки K .

Також, ключовим випадком буде ситуація, коли $\varphi|_{CD} = \varphi|_{MN} = \varphi^\circ = \varphi^* = 1$. У цьому ж випадку в процесі розв'язання відповідної задачі визначається положення точки K . Таким чином, в залежності від значень потенціалу φ° на додаткових еквіпотенціальних лініях (джерелах збурення MN) при збільшенні кількості еквіпотенціальних ділянок границі області, а також її зв'язності, область комплексного потенціалу $G_\omega = G_\omega(\varphi^\circ)$ будується індивідуально. У більш ширшому сенсі, в залежності від значення потенціалу $-\infty < \varphi^\circ < +\infty$, можливі різні випадки формування течії у фізичній області, а отже – побудови області квазікомплексного потенціалу [1].

Запропонована методика, окрім знаходження характеристичної функції течії та комплексного потенціалу, дозволяє одночасно будувати в заданій області динамічну сітку та моделювати лінії розділу фільтраційних потоків, розрахувати поле швидкості фільтрації, а також розраховувати обчислювальні (ітераційні) характеристики та досліджувати і аналізувати процеси їх збіжності і стійкості.

Розглянута методика наближення розв'язків крайових еліптичних задач, переноситься і на крайові задачі, які зводяться до конформних відображень довільної криволінійної області, обмеженої лініями течії та еквіпотенціальними лініями на область комплексного потенціалу, що має вигляд многокутника, границя якого складається із вертикальних та горизонтальних відрізків, променів, із декількома невідомими при відповідності вершин. Проблема існування розв'язків такого роду задач є давно розв'язаною лише в загальному. В кожному конкретному випадку розв'язність задачі забезпечується шляхом підбору відповідних параметрів.

VII. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У перспективі – моделювання та прогнозування роботи слабопровідних (близьких до сланцевих) пластів в умовах гідророзриву, а також – побудова ліній розділу різнокольорових рідин при прогнозуванні квазіідеальної течії у водоймах, обмежених лініями течії та еквіпотенціальними лініями.

VIII. ВИСНОВКИ

Чисельний метод конформних відображень використано для побудови динамічної сітки потенційних та квазіпотенційних полів, побудови різного роду профілів (зокрема – кривих розділу фільтраційних потоків) та поля швидкості із паралельним розрахунком різного роду характеристик (потенціалів, витрат, перетоків і т.ін.).

Розроблена методика розв'язання нелінійних крайових задач дозволяє розраховувати координати вузлів гідродинамічної сітки, обчислювати фільтраційні витрати, величини швидкості руху та інші параметри досліджуваних процесів, а також моделювати криві розділу фільтраційних потоків.

ЛІТЕРАТУРА

- [1] А.Я.Бомба, С.С.Каштан, Д.О.Пригорницький та С.В.Ярошак, *Методи комплексного аналізу*. Рівне, Україна: НУВГП, 2013. 415с.
- [2] A.Y.Bomba, S.S.Kashtan and V.V.Skopetskii, "Nonlinear Inverse Boundary-Value Problems of Conformal Mapping with a Controlling Potential," *Cybernetics and Systems Analysis*, Vol. 40, № 1, p. 58-65, 2004.
- [3] А.Я.Бомба та С.С.Каштан, "Нелінійні обернення крайових задач на конформні відображення з потенціалом керування," *Математичні методи та фізико-механічні поля*, т. 45, № 3, с. 69-76, 2002.
- [4] С.С.Каштан, "Математичне моделювання ідеальних та квазіідеальних полів при наявності джерела поперечних збурень," у *Інформаційні технології і автоматизація – 2022: матеріали XV Міжнародної науково-практичної конференції*, Одеса, Україна, 20-21 жовтня 2022. Одеса, Україна: ОНТУ, 2022. с. 27-30.
- [5] J.M.Ortega and W.C.Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York and London: Academic Press, 1970. 555p.
- [6] С.С.Каштан та М.В.Бойчура, "Числові методи комплексного аналізу моделювання повільного руху рідин у водоймах за умов керування," у *Матеріали Міжнародної наукової конференції «Сучасні проблеми математичного моделювання та обчислювальних методів»*, Рівне, Україна, 19-22 лютого 2015. Рівне, Україна: РВЦ РДГУ, 2015, с. 83-84.

Підвищення надійності системи логування

Іван Савка, Тарас Іванишин

*Кафедра інформаційних технологій
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна*

Анотація—В даному дослідженні проаналізовано аспекти надійності програмного забезпечення, зокрема у системах логування. Запропоновано архітектуру системи, яка дозволяє збільшити стійкість та доступність процесу логування.

Ключові слова—Програмне забезпечення, Логування, Надійність, Стійкість, Доступність.

I. ВСТУП

В епоху цифрової трансформації, коли дані мають велике значення, здатність надійно і ефективно обробляти великі обсяги інформації стає критично важливою. Одним з ключових аспектів цього процесу є логування, яке дозволяє системам відстежувати та аналізувати свою діяльність. Однак, зі збільшенням навантаження на системи, традиційні методи логування можуть стати недостатніми, що призводить до втрати важливої інформації та зниження продуктивності [1]. Тому збільшення надійності логування в розподілених системах високого навантаження стає актуальною та нагальною проблемою. У даній роботі розглянуто як сучасні підходи, так і технології Graylog Forwarder, які можуть допомогти вирішити цю проблему, зокрема, розподілене логування, використання системи управління базами даних в реальному часі та інші. На основі цього запропоновано шляхи покращення надійності логування в системах високого навантаження, а також наведено можливі виклики та перешкоди.

II. АНАЛІЗ ОСТАННІХ ПУБЛІКАЦІЙ

Надійність у програмах, що працюють у хмарних платформах – це міра ймовірності того, що хмарна система надає послуги, для яких вона була розроблена, у будь-який певний момент часу. Це означає, що сервіс має бути доступним, функціональним та продуктивним навіть у разі виникнення несподіваних несправностей у системі [2, 3].

Серед основних аспектів надійності варто виділити стійкість та доступність. Стійкість визначає методи та підходи для уникнення несправностей та відновлення після їх виникнення, а доступність групує методи для підходи для забезпечення постійного доступу до сервісів.

Логування – це процес створення, запису та обробки записів щодо різноманітні події та процеси, які відбуваються всередині ПЗ. Загалом основною аудиторією цих записів є інженери та системні адміністратори [4]. Кожне повідомлення має рівень, який репрезентує важливість повідомлення, рівень може означати як просте діагностичне повідомлення, так і повідомлення про критичний збій в роботі ПЗ. Зазвичай в процесі логування створюються повідомлення, що містять наступну інформацію, але не обмежуються нею:

- дата створення повідомлення;
- рівень повідомлення;
- джерело;
- текстовий опис.

Оскільки кожне повідомлення може потенційно містити критично важливу інформацію, то необхідно забезпечити його гарантоване відправлення і обробку [5].

У роботі [6] запропоновано підхід, що вводить поняття агента, який виконує обробку вхідних повідомлень та відправляє їх у віддалене сховище. Цей підхід використовує транспортний протокол для оркестрації збору, транспортування та зберігання даних журналу.

Це гарантує, що всі компоненти системи працюють разом безперебійно та можуть бути контрольовані застосунками оркестрації.

У технології Graylog Forwarder [9] використовуються підходи, які полягають в розгортанні проміжних сервісів, завданням яких є гарантія доставки повідомлення в кінцеву систему для обробки та збереження. Наведений далі підхід дозволяє збільшити надійність системи шляхом використання надлишкових сховищ та асинхронної відправки повідомлень.

III. МЕТОДИ

В даній роботі пропонується покращити та збільшити стійкість та доступність системи логування шляхом створення підсистеми, що буде слугувати проксі між клієнтами та сервером, який буде обробляти та зберігати повідомлення. Його роль полягатиме в надійному проміжному збереженні та відправленні повідомлень в кінцеву систему. Архітектура системи передбачає високу надійність та доступність за рахунок простоти та використання підходів для покращення цих якостей в системі. Архітектура системи зображена на Рисунку 1.

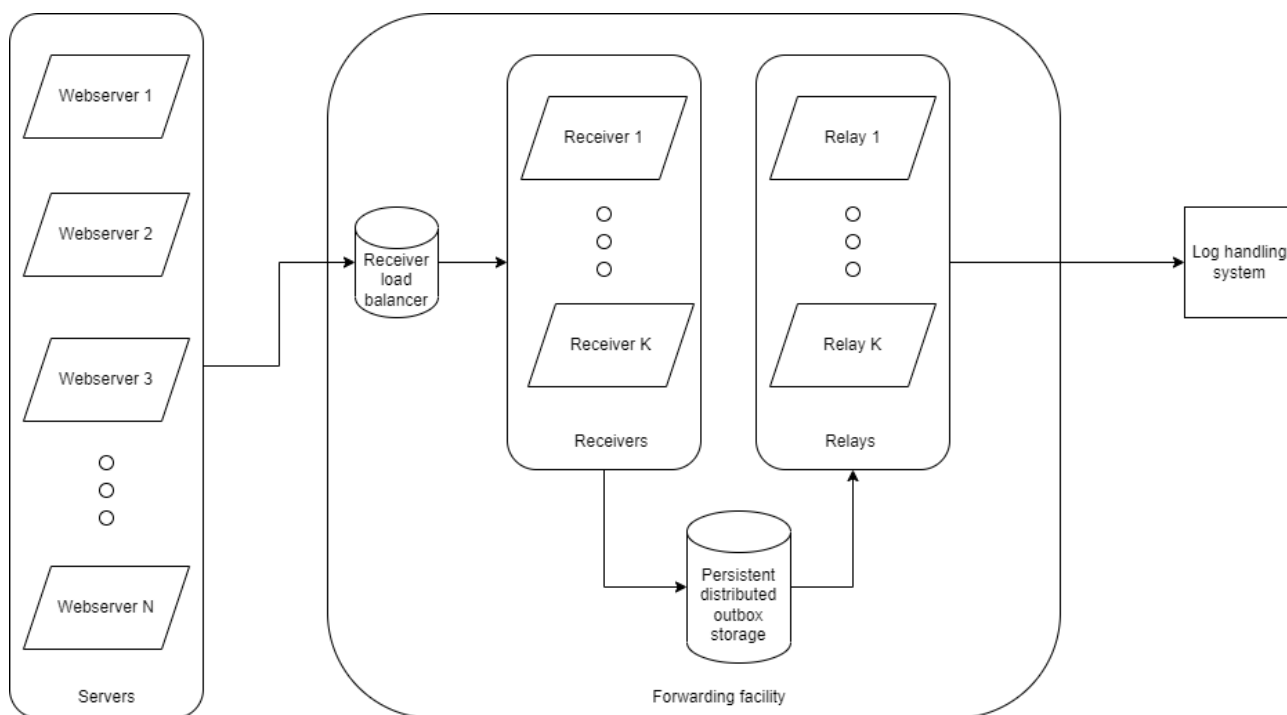


Рисунок 1 – Архітектура проксі-системи для надійного збереження та відправки повідомлень.

Даний підхід використовує прості Web-API сервіси, щоб записати отримане повідомлення у високодоступне сховище для даних, що забезпечує їх високу доступність у випадку критичних несправностей інфраструктури. Також, у випадку якщо високодоступне сховище недоступне, запис відбувається в резервне локальне сховище, після чого переноситься у головне сховище, як тільки воно відновить роботу. Після запису повідомлення зчитуються та відправляються на агенти, єдиною задачею яких є відправити ці повідомлення в систему призначення [7]. У випадку, якщо кінцева система є недоступною, агенти зазнають помилки, але жодне повідомлення не втрачається, процес доставки буде відновлений як тільки кінцева система відновить свою доступність.

На цій схемі отримувач (Receiver) отримує логи через певний мережний протокол та здійснює спробу записати їх в розподілене сховище. Якщо спроба виявляється невдалою у зв'язку з несправностями сховища, то запис відбувається в локальне сховище, яким може виступати файл або інший сервіс, що працює в тому контексті, що й отримувачі. Як тільки

сховище стає доступним знову, отримувачі здійснюють спробу зчитати повідомлення з резервного сховища/сховищ і зберегти в основне. Відправник (Relay) здійснює обробку повідомлень, що були записані в основне сховище. Обробка полягає у відправці повідомлення в кінцеву систему для обробки. Прикладом розподіленого сховища може виступати Azure Cosmos DB. Це розподілена база даних, яка гарантує високий SLA 99.99%. Також ця система БД має дуже корисну функцію під назвою Change Feed, що дозволяє уникнути використання брокера повідомлень при реалізації описаної архітектури [10].

При використанні геореплікованого основного сховища можна досягти високої доступності системи логування. Зокрема, можливо буде реплікувати систему між різними геовіддаленими регіонами, що значно збільшить надійність, оскільки кожен екземпляр системи буде незалежним один від одного, а також значно збільшить доступність, оскільки зменшиться час відгуку навколо точок геореплікації [8].

Таким чином, описаний підхід визначає систему для проксування повідомлень, що гарантує доставку повідомлень, незалежно від того, чи кінцева система справно функціонує.

IV. РЕЗУЛЬТАТИ

В даній роботі було проаналізовано аспекти надійності в розподілених хмарних системах, зокрема у системі логування. Описано новий підхід та запропоновано архітектуру системи, яка дозволяє збільшити стійкість та доступність процесу логування.

Важливо зауважити, що імплементація цього підходу має певні складнощі, з яких можна виділити комплексність системи, яка ускладнює її розгортання та підтримку, та відносно високу ціну інфраструктури, яка частково компенсується шляхом спільного використання ресурсів.

V. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

При подальших дослідженнях, покращення підходу може підвищити його цінову ефективність, а також призвести до подальшого збільшення надійності.

VI. ВИСНОВКИ

Надійність є критичним аспектом для будь-якої розподіленої системи, оскільки вона репрезентує її роботоздатність. А логування станів системи є також важливою, адже саме вона є першим джерелом інформації про несправності у роботі системи. Ключовою перевагою застосування стійкої і доступної системи логування є гарантія кінцевої консистентності інформації, яка може бути критично важливою для клієнта та його команди.

ЛІТЕРАТУРА

- [1] Ian Giorton, *Foundations of Scalable Systems*, Sebastopol, CA: O'Reilly Media, 2022, 340 p.
- [2] Zimba Yolam, "Building Reliable Cloud Systems through Chaos Engineering," *International Journal of Managing Information Technology*, vol. 14, no. 2, pp. 1-7, 2022, doi: 10.5121/ijmit.2022.14201
- [3] M.A. Shahid, M.M. Alam and M.M. Su'ud, "Achieving Reliability in Cloud Computing by a Novel Hybrid Approach," *Sensors*, vol. 23, p. 1965, 2023, doi: 10.3390/s23041965
- [4] Sina Gholamian and Paul A. S. Ward, "A Comprehensive Survey of Logging in Software: From Logging Statements Automation to Log Mining and Analysis," *ArXiv*, 2021, Accessed: Nov.12, 2023. [Online]. Available: <https://arxiv.org/pdf/2110.12489.pdf>.
- [5] Q. Cheng, A. Saha, W. Yang, C. Liu, D. Sahoo and S. Hoi, "LogAI: A Library for Log Analytics and Intelligence," *ArXiv*, 2023, Accessed: Nov. 12, 2023. [Online]. Available: <https://arxiv.org/pdf/2301.13415.pdf>.
- [6] Mayol Arnao Reinaldo, Nuñez Luis A. and Lobo Antonio, "An Approach to Log Management: Prototyping a Design of Agent for Log Harvesting," *ArXiv*, 2011, Accessed: Nov. 12, 2023. [Online]. Available: <https://arxiv.org/ftp/arxiv/papers/1112/1112.0795.pdf>.
- [7] Pattern: Transactional outbox, [Online]. Available: <https://microservices.io/patterns/data/transactional-outbox.html>.
- [8] Multi-region load balancing with Traffic Manager, Azure Firewall, and Application Gateway, [Online]. Available: <https://learn.microsoft.com/en-us/azure/architecture/high-availability/reference-architecture-traffic-manager-application-gateway>.

- [9] Forwarder, [Online]. Available: https://go2docs.graylog.org/5-0/getting_in_log_data/forwarder.html.
- [10] Achieve high availability with Azure Cosmos DB, [Online]. Available: <https://learn.microsoft.com/en-us/azure/cosmos-db/high-availability>.

Монотонні предикати в теорії ігор

Оксана Микицей

Кафедра алгебри та геометрії
Прикарпатський національний університет імені Василя Стефаника
Івано-Франківськ, Україна

Анотація – Монотонні предикати застосовано у ситуаціях неповної або неточної інформації, наприклад, у іграх двох гравців. Опис неповної інформації про позицію гравця у грі за допомогою неперервних напівграток. Отримано апроксимативну теорему про мінімакс.

Ключові слова: Монотонний предикат, Неперервна напівгратка, Ігри з неповною інформацією.

I. ВСТУП

Метою роботи є застосування стандартного інструментарію денотаційної семантики мов програмування, а саме монотонних предикатів на неперервних напівгратках, до теорії ігор. Це майже немає сенсу для ігор з повною інформацією, але все стає інакше, якщо гравці мають лише часткову інформацію про фактичні позиції у грі і, крім того, мають обмежений контроль над змінами цих позицій.

II. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Е. Дійкстра в своїй праці [1] зауважив, що зміст кожної програми визначається тим, як істинність припущень (предикатів) про вхідні дані впливає на істинність припущень (предикатів) про результат її роботи. Всі можливі (ймовірно, неповні) "порції інформації" про дані чи стан системи у деякий момент утворюють область обчислення. На цій множині є природний частковий порядок, який відбиває ієрархію інформації чи знання: що більше інформації несе елемент (що більш обмежувачим/специфічним він є), то більшим ми його вважаємо. У [2] також обґрунтовано, чому природно вимагати, щоб область обчислення була неперервною напівграткою у сенсі теорії областей, яку створив Дана Скотт [3]. Відповідно, функцію, яка кожній "порції інформації" зіставляє її "ступінь правдивості", ми, слідуючи [4], називаємо монотонним предикатом на області обчислення. Монотонні предикати на неперервних напівгратках зі значеннями у цілком дистрибутивних гратках неминуче виникають у цьому контексті.

III. РЕЗУЛЬТАТИ

Нехай S – неперервна напівгратка з нижнім елементом, яку можна вважати множиною всіх порцій неповної інформації, яку можна отримати в певний момент часу. Множину всіх неперервних за Скоттом відображень з напівгратки S в компактну гаусдорфову гратку Лоусона L позначимо $[S \rightarrow L]$. Елементи $[S \rightarrow L^{op}]^{op}$ [4] називаємо L -нечіткими монотонними предикатами на S .

Множину всіх L -нечітких нормованих монотонних предикатів на неперервній напівгратці S позначимо $M_{[L]}S = [S \rightarrow L^{op}]^{op}$.

Розглянемо грубі ігри двох гравців у розширеній формі, коли перший і другий гравці ходять по черзі, а результат гри є елементом цілком дистрибутивної гратки L .

Монотонний предикат $R_i(x_i) = m_{x_i R_i} \in M_{[L]}S_{-i}$ повністю описує позицію у грі з точки зору другого гравця після ходу $x_i \in S_i$ першого гравця. І навпаки, $R_{-i}(x_{-i}) = m_{x_{-i} R_{-i}} \in M_{[L]}S_{i+1}$ зображує позицію першого гравця після ходу x_{-i} другого.

Побудовано монотонні предикати виплат $\bar{w}_i \in M_{[L]}S_i$ і $\bar{w}_{-i} \in M_{[L]}S_{-i}$:

$$\bar{w}_i(x_i) = \sup\{w_i(x)|x \in S_i, x \ll x_i\} \text{ і } \bar{w}_{-i}(x_{-i}) = \text{s}\ddot{\text{u}}\text{p}\{w_{-i}(x)|x \in S_{-i}, x \ll x_{-i}\},$$

що описують нижні обмеження на результати при правильній грі.

Знайдено рекурентні співвідношення, які дозволяють обчислити ці предикати, рухаючись від останнього можливого ходу гри.

Теорема. Для скінченної послідовності $S_1, S_{-1}, S_2, S_{-2}, \dots$ неперервних напівграток з нулями і грубої гри двох гравців у розширеній формі, у якій початковий стан задається монотонним предикатом $m_0 \in M_{[L]}S_1$, а ходи гравців – неоднозначними зображеннями $R_i : S_i \Rightarrow^L S_{-i}$ і $R_{-i} : S_{-i} \Rightarrow^L S_{i+1}$, предикати виплат задовольняють рекурсивні співвідношення

$$\bar{w}_i(x_i) = \sup_{x \ll x_i} \text{inf}\{R_i(x)(x_{-i}) \vee \bar{w}_{-i}(x_{-i}) | x_{-i} \in S_{-i}\}$$

і

$$\bar{w}_{-i}(x_{-i}) = \text{s}\ddot{\text{u}}\text{p}_{x \ll x_{-i}} \text{inf}\{R_{-i}(x)(x_{i+1}) \vee w_{i+1}(x_{i+1}) | x_{i+1} \in S_{i+1}\}.$$

Зауважимо, що супремуми та інфімуми у теоремі не обов'язково досяжні. Отже, отримано "апроксимативну теорему про мінімакс": *Для грубої скінченної дійснозначної гри розширеної форми з нульовою сумою, найменшою верхньою гранню мінімального виграшу першого гравця є найбільша нижня грань максимального програшу другого гравця.*

IV. ОБГОВОРЕННЯ

Розглянемо як застосовувати описані вище функції, щоб отримати у грі кращі результати. Для спрощення припустимо, що гратка L лінійно впорядкована, наприклад, $L = [0; 1]$ і $\bar{\omega}_0(1) = \alpha$. Перший гравець може виявитися нездатним отримати α , тому оберемо як завгодно близьке $\alpha' < \alpha$. Оскільки $\bar{\omega}_0(1) > \alpha'$, існує таке $x_1 \in S_1$, що $m_0(x_1) \wedge \omega_1(x_1) > \alpha'$. Нехай перший гравець зробить свій вибір x_1 . Якщо другий гравець погодиться на $m_0(x_1) > \alpha'$ і вийде з гри, то перший отримає не менше ніж α' . Інакше за цим слідує хід $x_{-1} \in S_{-1}$, де $\omega_1(x_1) > \alpha'$ і, дещо послабивши x_1 до x , можемо розраховувати на $R_1(x)(x_{-1}) \vee \bar{\omega}_{-1}(x_{-1}) > \alpha'$. Якщо $R_1(x)(x_{-1}) > \alpha'$, то перший гравець може вийти з гри з результатом $> \alpha$, або, завдяки $\bar{\omega}_{-1}(x_{-1}) > \alpha'$, він обере таке $x_2 \in S_2$, що $R_{-1}(x_{-1})(x_2) \wedge \bar{\omega}_2(x_2) > \alpha'$, і т.д. Так перший гравець може гарантувати довільний результат $\alpha' < \alpha$.

З іншого боку, з $\bar{\omega}_0(1) = \alpha$ впливає $m_0(x_1) \wedge \bar{\omega}_1(x_1) \leq \alpha$ для всіх ходів $x_1 \in S_1$. Отже, після кожного ходу другий гравець може або отримати $m_0(x_1) \leq \alpha$, або скористатись з $\omega_1(x_1) \leq \alpha$, щоб знайти для кожного $\alpha'' > \alpha$ такий хід $x_{-1} \in S_{-1}$, що $R_1(x)(x_{-1}) \vee \omega_{-1}(x_{-1}) < \alpha''$. Після цього ходу перший гравець може або взяти $R_1(x)(x_{-1}) \vee \omega_{-1}(x_{-1}) < \alpha''$ і вийти, або продовжити гру, і, оскільки $\omega_{-1}(x_{-1}) < \alpha''$, в обох випадках результат другого гравця буде не гіршим від $\alpha'' > \alpha$.

V. ВИСНОВКИ

Результати роботи є внеском у теорію неперервних областей і відкривають можливості застосування граткозначних монотонних предикатів у моделюванні прийняття рішень в умовах невизначеності, зокрема, в умовах ігор з неповною інформацією.

VI. ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Перспективи подальших досліджень пов'язані з послабленням деяких вимог до напівґраток та їх відображень, а також з глибшим аналізом монотонних предикатів з погляду і засобами теорії категорій.

ЛІТЕРАТУРА

- [1] E. W. Dijkstra, “Guarded commands, nondeterminacy and formal derivation of programs”, *Communications of the ACM*, vol. 18, no. 8, pp. 453–457, 1975, doi: 10.1145/360933.360975.
- [2] A. Edalat, “Domains for Computation in Mathematics, Physics and Exact Real Arithmetic”, *Bulletin of Symbolic Logic*, vol. 3, no. 4, pp. 401–452, 1997, doi: 10.2307/421098.
- [3] J. D. Lawson, M. Mislove, G. Gierz, K.H. Hofmann, K. Keimel, D. S. Scott, “Continuous Lattices and Domains”, *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, vol. 93, 2003.
- [4] R. Heckmann, M. Huth, “A duality theory for quantitative semantics”, *Proceedings of the 11th International Workshop on Computer Science Logic*, Springer Verlag, pp. 255 – 274, 1998.
- [5] O. Nykyforchyn, O. Mykitysey, “Ambiguous Representations of Semilattices, Imperfect Information, and Predicate Transformers”, *Order*, vol. 37, no. 2, pp. 319–339, 2019, doi: 10.1007/s11083-019-09508-0.
- [6] O. Nykyforchyn, D. Repovš, “Ambiguous representations as fuzzy relations between \mathbf{L} sets”, *Fuzzy Sets and Systems*, vol. 173, no.1, pp. 25–44, 2011, doi: 10.1016/j.fss.2011.02.007.


Weak sinusoidal signal extraction from white noise using convolutional neural network

Mykola Kozlenko

Data Science Group

SoftServe Inc

Austin, TX, US

 <https://orcid.org/0000-0002-2502-2447>

Abstract—A great number of analog and digital data communications schemes use the sinusoidal waveform as a basic elementary signal, including the spread spectrum data exchange techniques. Detection of the presence of the sinusoidal waveform in a mixture of signal and noise is a common task, regardless the specific modulation scheme. This paper presents the machine learning-based approach for detection of the sinusoidal wave. It presents the structure of the convolutional neural network, as well as the performance metrics for the sinusoidal signals detection. The paper provides an assessment of the overall accuracy for the binary signals. It reports the overall accuracy value of 0.93 for the sinusoidal signal detection in the presence of additive white Gaussian noise at the signal-to-noise ratio value of -20 dB for a balanced dataset.

Keywords—digital communications, modulation, manipulation keying, demodulation, detection, bit error rate, machine learning, deep learning, convolutional neural network, JT65

I. INTRODUCTION

The most fundamental digital modulation techniques are based on the use of sinusoidal waveform as a carrier [1], [2]. Automatic information extraction in signal monitoring requires some intermediate steps such as modulation classification, signal identification [3], etc. Demodulation includes the estimation of sinusoidal waveform parameters, such as amplitude, phase, and frequency. This paper presents the method for detection of the presence of the sinusoidal component in the mixture of signal and noise. It can be implemented in the software-defined radios, cognitive radios, and wireless sensor networks. The presented method of the detection of the continuous sinusoidal wave is also applicable in radiotelegraphy as well as in radiolocation.

II. RELATED WORK

Paper [4] presents a novel signal extraction algorithm from chaotic background using wavelet packet transform. In [5], the authors present a new method for the detection of signals in "noise", which is based on the premise that the "noise" is chaotic with at least one positive Lyapunov exponent. A novel method of detection, extraction, and estimation of amplitude, phase and frequency of sinusoids of time-varying nature is presented in [6]. In the paper [7], the problem of non-coherent detection of a sinusoidal carrier is considered in the presence of Gaussian noise. The author of [8] investigates the convexity properties of error probability in the detection of binary-valued scalar signals corrupted by additive noise. The paper [9] solves the problem of weak radio signals demodulation for Frequency Shift Keying (FSK) and Phase Shift Keying (PSK) signals with Machine Learning (ML) using perceptron model with randomization. Paper [10] provides the bit-error-rate performance of the spread spectrum binary communication system with noise shift keying scheme using the entropy demodulation. Reference [11] presents the idea of using CNNs with deep learning structure to predict future symbols based on the received signal, to further reduce inter-symbol interference and to obtain a better Bit Error Rate (BER) performance in chaotic baseband wireless communication systems.

III. METHODOLOGY

Fig. 1 shows the time domain waveform and the spectrogram of the mixture of signal and Additive White Gaussian Noise (AWGN) at SNR -20 dB.

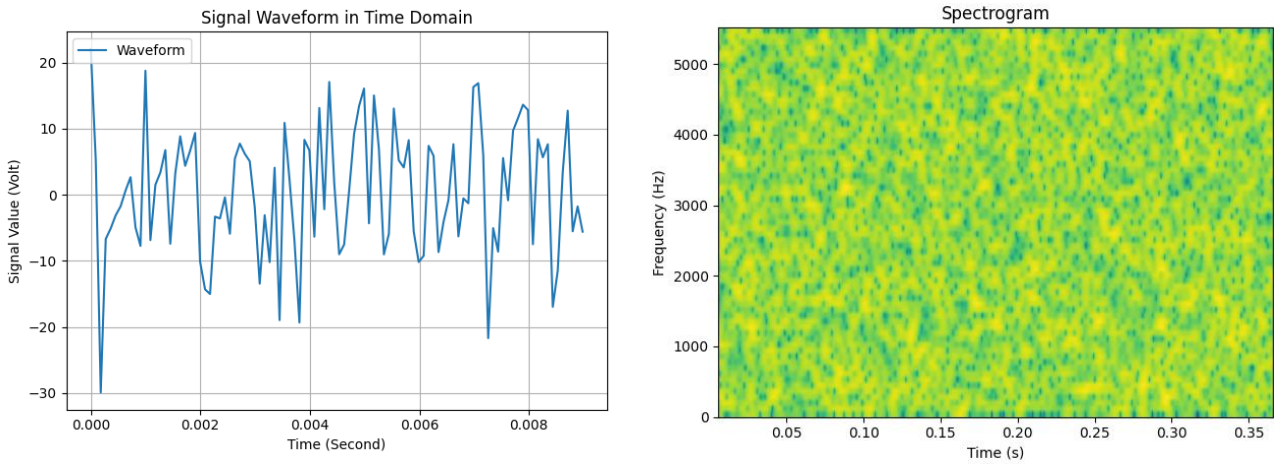


Fig. 1. Time domain waveform and the spectrogram of the mixture of signal and AWGN

Fig. 2 shows the structure of the convolutional neural network that is used for the detection.

Layer (type)	Output Shape	Param #
reshape (Reshape)	(None, 4096, 1)	0
batch_normalization	(None, 4096, 1)	4
conv1d (Conv1D)	(None, 4096, 128)	2176
max_pooling1d	(None, 2048, 128)	0
dropout (Dropout)	(None, 2048, 128)	0
batch_normalization_1	(None, 2048, 128)	512
conv1d_1 (Conv1D)	(None, 2048, 64)	131136
max_pooling1d_1	(None, 1024, 64)	0
dropout_1 (Dropout)	(None, 1024, 64)	0
batch_normalization_2	(None, 1024, 64)	256
conv1d_2 (Conv1D)	(None, 1024, 32)	32800
max_pooling1d_2	(None, 512, 32)	0
dropout_2 (Dropout)	(None, 512, 32)	0
flatten (Flatten)	(None, 16384)	0
batch_normalization_3	(None, 16384)	65536
dense (Dense)	(None, 64)	1048640
batch_normalization_4	(None, 64)	256
dense_1 (Dense)	(None, 2)	130

=====
Total params: 1281446 (4.89 MB)
Trainable params: 1248164 (4.76 MB)
Non-trainable params: 33282 (130.01 KB)
=====

Fig. 2. Structure of the convolutional neural network

In this research, the modulation scheme is On-Off-Keying (OOK). The main baseband frequency is 1270.5 Hz. The baseband bandwidth is 5,512.5 Hz. The sampling rate is 11,025.0 Hz. The bit duration time is 0.3715 s, 4096 samples (exactly as it is in JT65 protocol [12, 13]). The bit rate is 2.6917 bit/s. The signal-to-noise ratio (SNR) is -20 dB, the normalized signal-to-noise ratio (E_b/N_0) is $+13.1$ dB, and the spreading factor is $+33.1$ dB). The size of the training dataset is 10720 records, the validation set size is 2680, and the test set size is 6600 records. Initial training/test split

ratio is 0.33; training/validation split ratio is 0.2. The loss function: sparse categorical cross-entropy, optimizer: adam. Two separate output neurons are used in the model. There is no need for one-hot encoding for labels due to the sparse loss function. The binary model can be easily converted to a multiclass classifier for m-ary signals.

IV. RESULTS

The classification report for the test data is shown in Table 1, the confusion matrix - in Table 2.

TABLE 1. CLASSIFICATION REPORT

Class	Classification Metric			
	Precision	Recall	F1-Score	Support
0 – "space"	0.89	0.98	0.93	3351
1 – "mark"	0.97	0.87	0.92	3249
Accuracy			0.93	6600
Macro avg	0.93	0.93	0.93	6600
Weighted avg	0.93	0.93	0.93	6600

TABLE 2. CONFUSION MATRIX

True	Predicted		
	0	1	All
0	3273	78	3351
1	410	2839	3249
All	3683	2917	6600

V. DISCUSSION

The purpose of this study was to gain a better understanding of the ability of machine learning algorithms to detect the baseband sinusoidal signals. The result of the present study supports the hypothesis that the presence of sinusoidal signal can be detected with ML techniques. The results of this research provide supporting evidence that it is possible even in the presence of noise. This is the main take away from this paper. This pattern of results is consistent with our previous works those deal with ML-based demodulation for frequency shift keying modulation scheme. These results represent the direct demonstration of sinusoidal signal detection within the scope of the ML approach. There are at least three potential limitations concerning the results of this study. A first limitation is that we used signals generated with the same power. A second potential limitation is that we used supervised machine learning approach. This means that the algorithm should be trained on previously known data. Unsupervised approach looks like more promising technique. The third limitation is that we considered only the case of AWGN interference. Despite these limitations, the results suggest practical implication that sinusoidal signals can be detected using ML.

VI. FUTURE RESEARCH

In terms of future research, it would be useful to use the quantum optimization algorithms [14]. The results might be applied in data communication applications in mobile robotics [15]. Also, it would be useful to research the ability of unsupervised ML [16] approach in the context of this task. Other types of interferences (chaotic [4, 5], intermodulation distortion [17], etc.) should be studied as well.

VII. CONCLUSION

The main conclusion that can be drawn is that sinusoidal baseband signals can be detected using supervised machine learning approach. In summary, this paper presents the values of quality metrics. The overall accuracy 0.93 and minimum F1-score is 0.92 at SNR value of -20 dB.

VIII. ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of scientists of the Department of Information Technology of Vasyl Stefanyk Precarpathian National University for scientific guidance given in discussions and technical assistance helped in the actual research.

IX. ETHICS DECLARATIONS

The author has nothing to disclose.

REFERENCES

- [1] H. Wu and X. Hua, "A review of digital signal modulation methods based on wavelet transform," *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, Beijing, China, 2020, pp. 1123-1128, doi: 10.1109/ICMA49215.2020.9233682
- [2] S. Nagul, "A review on 5G modulation schemes and their comparisons for future wireless communications," *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, Vijayawada, India, 2018, pp. 72-76, doi: 10.1109/SPACES.2018.8316319
- [3] Q. Shi and Y. Karasawa, "Maximum likelihood based modulation classification for unsynchronized QAMs," *2008 IEEE Global Telecommunications Conference*, New Orleans, LA, USA, 2008, pp. 1-5, doi: 10.1109/GLOCOM.2008.ECP.664
- [4] Kaiyan Zhu and Shuxun Wang, "Sinusoidal signal extraction from chaotic background using wavelet packet transform," *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004.*, Beijing, China, 2004, pp. 176-179 vol.1, doi: 10.1109/ICOSP.2004.1452610
- [5] S. Haykin and Xiao Bo Li, "Detection of signals in chaos," in *Proceedings of the IEEE*, vol. 83, no. 1, pp. 95-122, Jan. 1995, doi: 10.1109/5.362751.
- [6] A. K. Ziarani, I. M. Blumenfeld, and A. Konrad, "Experimental verification of a novel method of extraction of nonstationary sinusoids," *The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.*, Tulsa, OK, USA, 2002, pp. I-455, doi: 10.1109/MWSCAS.2002.1187256
- [7] C. Ozturk, B. Dulek and S. Gezici, "Convexity properties of detection probability for noncoherent detection of a modulated sinusoidal carrier," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12410-12415, Dec. 2018, doi: 10.1109/TVT.2018.2876516.
- [8] M. Azizoglu, "Convexity properties in binary detection problems," in *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1316-1321, July 1996, doi: 10.1109/18.508867
- [9] I. Lazarovych *et al.*, "Software implemented enhanced efficiency BPSK demodulator based on perceptron model with randomization," *2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2021, pp. 221-225, doi: 10.1109/UKRCON53503.2021.9575458
- [10] M. Kozlenko and A. Bosyi, "Performance of spread spectrum system with noise shift keying using entropy demodulation," *2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, 2018, pp. 330-333, doi: 10.1109/TCSET.2018.8336213
- [11] Hai-Peng Ren, Hui-Ping Yin, Hong-Er Zhao, C. Bai, C. Grebogi, "Artificial intelligence enhances the performance of chaotic baseband wireless communication," *IET Communication*, vol. 15, 2021, pp. 1467-1479, doi: 10.1049/cmu2.12162
- [12] J. Taylor, "The JT65 Communications Protocol," *QEX*, Sept.-Oct. 2005, pp. 3-12
- [13] J. Taylor, "WSJT: New Software for VHF Meteor-Scatter Communication," *QST*, Dec. 2001, pp. 36-41
- [14] V. Tkachuk and M. Kozlenko, "Improved quantum genetic algorithm on multilevel quantum systems for 0-1 knapsack problem," in *Advances in Artificial Systems for Logistics Engineering (Lecture Notes on Data Engineering and Communications Technologies, vol. 135)*, Z. Hu, Q. Zhang, S. Petoukhov, and M. He, Eds., Kyiv, Ukraine, Feb. 20-22, 2022, pp. 51-70, doi: 10.1007/978-3-031-04809-8_5
- [15] V. Tkachuk, M. Kozlenko, M. Kuz, I. Lazarovych, and M. Dutchak, "Function optimization based on higher-order quantum genetic algorithm," *Electronic Modeling*, vol. 41, no. 3, pp. 43-58, 2019, doi: 10.15407/emodel.41.03.043
- [16] D. Boulinguez, C. Garnier, Y. Delignon, and L. Clavier, "Unsupervised characterization of digital modulations," *6th International Conference on Signal Processing, 2002.*, Beijing, China, 2002, pp. 1316-1319 vol.2, doi: 10.1109/ICOSP.2002.1180034
- [17] R. Figueiredo, N. B. Carvalho, A. Piacibello and V. Camarchia, "Nonlinear dynamic RF system characterization: envelope intermodulation distortion profiles - a noise power ratio-based approach," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 9, pp. 4256-4271, Sept. 2021, doi: 10.1109/TMTT.2021.3092398

Scientific electronic edition

2023 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE)

Conference Proceedings

November 29-30, 2022
Ivano-Frankivsk, Ukraine

Issue editors: Mykola Kuz and Mykola Kozlenko

Publication language: English
(includes papers in English and Ukrainian)

First Online: Nov. 29, 2023

Publisher:

Vasyl Stefanyk Precarpathian National University
1 Stepana Bandery Street
Ivano-Frankivsk, 76018, Ukraine
E-mail: vdvcit@pnu.edu.ua
Tel: +380 (342) 75-13-08

Publisher certificate DK #7616 2022-05-26

Видавець

Прикарпатський національний університет
імені Василя Стефаника
76018, м. Івано-Франківськ,
вул. С. Бандери, 1, тел.: +380 (342) 75-13-08
E-mail: vdvcit@pnu.edu.ua

Свідоцтво суб'єкта видавничої справи ДК № 7616 від 26.05.2022

ISBN 978-966-640-549-7