



ISSN 2224-087X

ЕЛЕКТРОНІКА ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ELECTRONICS
AND INFORMATION TECHNOLOGIES

Збірник наукових праць

Випуск 29



2025

**ELECTRONICS
AND
INFORMATION
TECHNOLOGIES**

Issue 29

Scientific journal

Published 4 issue per year

Published since 1966

**ЕЛЕКТРОНІКА
ТА
ІНФОРМАЦІЙНІ
ТЕХНОЛОГІЇ**

Випуск 29

Збірник наукових праць

Виходить 4 рази на рік

Видається з 1966 р.

**Ivan Franko National
University of Lviv**

**Львівський національний
університет імені Івана Франка**

2025

ЗАСНОВНИК: ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Друкується за ухвалою Вченої Ради
Львівського національного університету
імені Івана Франка
протокол №80/03 від 28.03.2025 р.

У 1966–2010 рр. збірник виходив під назвою «Теоретична електротехніка»

Збірник «Електроніка та інформаційні технології» містить оригінальні результати досліджень з електронного матеріалознавства, моделювання фізичних явищ, процесів і систем електроніки, обробки сигналів і зображень, інформаційних технологій.

Редактори

Проф. *І. Карбовник* – головний співредактор
Проф. *І. Бордун* – головний співредактор
Проф. *О. Крупич* – відповідальний редактор
С.н.с. *Я. Шмигельський* – відповідальний секретар

Редакційна колегія

д-р техн. наук, проф. *О. Андрейків*
д-р габіл., проф. *Б. Андрієвський*
д-р фіз.-мат. наук, проф. *І. Болеста*
канд. фіз.-мат. наук, доц. *С. Вельгош*
д-р фіз.-мат. наук, проф. *П. Венгерський*
д-р техн. наук, проф. *Р. Воробель*
д-р фіз.-мат. наук, проф. *Р. Головчак*
д-р техн. наук, д-р габіл., проф. *Ф. Івацішин*
д-р фіз.-мат. наук, проф. *О. Кушнір*
д-р фіз.-мат. наук, проф. *А. Лучечко*
д-р техн. наук, проф. *Л. Муравський*
д-р техн. наук, проф. *М. Назаркевич*
д-р фіз.-мат. наук, проф. *І. Оленич*
д-р фіз.-мат. наук, проф. *Б. Павлик*
д-р техн. наук, доц. *Б. Павлишенко*
д-р техн. наук, проф. *С. Рендзіняк*
д-р фіз.-мат. наук, проф. *Б. Русин*
д-р фіз.-мат. наук, проф. *М. Пригула*
д-р габіл., проф. *Ц. Славінські*
канд. фіз.-мат. наук, доц. *Ю. Фургала*
д-р габіл., проф. *Б. Ціж*
д-р габіл., проф. *Бушта Шахраї*
д-р фіз.-мат. наук, проф. *Г. Шинкаренко*
канд. фіз.-мат. наук, доц. *Р. Шувар*
д-р фіз.-мат. наук, проф. *І. Яворський*

Адреса редакційної колегії:

Львівський національний університет імені Івана
Франка, факультет електроніки та інформаційних
технологій, вул. Ген. М. Тарнавського, 107,
79017, Львів, Україна
тел. (+38) (093) 864-01-19

e-mail: elit@lnu.edu.ua

web-сайт: <http://publications.lnu.edu.ua/collections/index.php/electronics/index>

Реєстрація суб'єкта у сфері друкованих медіа:

Рішення Національної ради України з питань
телебачення і радіомовлення № 1877 від
30.05.2024 р. Ідентифікатор медіа R30-04912

У 1966–2010 рр. збірник виходив під назвою «Теоретична електротехніка»

“Electronics and information technologies” journal contains original research results on electronics material science, modelling of physical phenomena, processes and electronic systems, signal and image processing and information technologies.

Editors

Prof. *I. Karbovnyk* – Chief Co-Editor
Prof. *I. Bordun* – Chief Co-Editor
Prof. *O. Krupych* – Managing Editor
Sen. Res. *Ya. Shmygelsky* – Technical Editor

Editorial Board

O. Andreykiv, Dr. Sc., Prof.
B. Andrievsky, Dr. Habil., Prof.
I. Bolesta, Dr. Sc., Prof.
S. Velgosh, PhD, Assoc. Prof.
P. Vengersky, Dr. Sc., Prof.
R. Vorobel, Dr. Sc., Prof.
R. Holovchak, Dr. Sc., Prof.
F. Ivaschyshyn, Dr. Sc., Dr. Habil., Prof.
O. Kushnir, Dr. Sc., Prof.
A. Luchechko, Dr. Sc., Prof.
L. Muravsky, Dr. Sc., Prof.
M. Nazarkevych, Dr. Sc., Prof.
I. Olenych, Dr. Sc., Prof.
B. Pavlyk, Dr. Sc., Prof.
B. Pavlyshenko, Dr. Sc., Assoc. Prof.
S. Rendzinyak, Dr. Sc., Prof.
B. Rusyn, Dr. Sc., Prof.
M. Prytula, Dr. Sc., Prof.
C. Slawinski, Dr. Habil., Prof.
Yu. Furgala, PhD, Assoc. Prof.
B. Tsizh, Dr. Habil., Prof.
Bouchta SAHRAOUI, Dr. Habil., Prof.
G. Shynkarenko, Dr. Sc., Prof.
R. Shuvar, PhD, Assoc. Prof.
I. Yavorsky, Dr. Sc., Dr. Habil., Prof.

Editorial office address:

Ivan Franko National University of L'viv,
Faculty of Electronics and Computer Technologies
107 Tamavsky St., UA–79017,
Lviv, Ukraine
tel. (+38) (093) 864-01-19

АДРЕСА РЕДАКЦІЇ, ВИДАВЦЯ І ВИГОТОВЛЮВАЧА:

Львівський національний університет імені Івана Франка
вул. Університетська, 1, 79000 Львів, Україна

Свідцтво про внесення суб'єкта видавничої справи до Державного реєстру видавців,
виготівників і розповсюджувачів видавничої продукції. Серія ДК № 3059 від 13.12.2007 р.

© Львівський національний університет імені Івана Франка, 2025

ЗМІСТ

ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ

Application of Artificial Neural Network and Fuzzy Inference in the Houseplant Watering Intelligent System. <i>Roman Korostenskyi, Igor Olenych</i> (8)	5
Властивості мереж, побудованих на цифрових зображеннях. <i>Олег Кушнір, Олексій Кравчук, Володимир Франів</i> (16)	13
Enhancement of Sensor Panel Tactile Touch Interface. <i>Oleksandr Karpin, Zinovii Liubun, Vasyl Mandziy, Oleh Tereshchuk, Nestor Hotsiy</i> (8)	29
Визначення найкращого оптимізатора для нейромережі при розробці систем автоматичної класифікації зображень. <i>Андріан Козинець</i> (10)	37
Ефективність обробки паралельних запитів у середовищі Amazon Web Services з використанням контейнерів. <i>Олег Сігунов, Лідія Демків</i> (10)	47
Deep Learning Objects Detection through Model Optimization to Accelerate the Process on General-purpose Single-board Platforms. <i>Dmytro Myroniuk, Bohdan Blagitko</i> (12)	57
Проблеми застосування нейронних мереж для прогнозування ціни віртуальних активів. <i>Андрій Цемко, Максим Мацьків</i> (10)	69
Energy conservation as one of the components of the management system for the smart sustainable workspaces. <i>Maksym Yakubovych, Vasyl Lyashkevych, Roman Shuvar</i> (16)	79

МОДЕЛЮВАННЯ ПРОЦЕСІВ ТА ЯВИЩ

The Inverse Gaussian Plume Method for Estimating the Level of Air Pollution. <i>Volodymyr Hura, Liubomyr Monastyrskiy</i> (16)	95
Метод оцінки доплерівського фактору широкосмугового сигналу в умовах дії мультиплікативної завади з невідомим розподілом ймовірності. <i>Павло Костенко, Микита Альонкін, Валерій Слободянюк, Михайло Ровецький, Іван Ровецький</i> (12)	111
Комп'ютеризовані оптичні експерименти з використанням портативних спектрометрів та інтеграцією з Amazon Web Services. <i>Андрій Крупич, Іван Карбовник</i> (12)	123

МАТЕРІАЛИ ЕЛЕКТРОННОЇ ТЕХНІКИ

Photosensitive Structures Based on the Silver-doped Porous Silicon and Reduced Graphene Oxide. <i>Igor Olenych, Lyubomyr Monastyrskii, Petro Parandiy, Roman Serkiz, Ivan Karbovnyk</i> ((10))	135
Зміни термостимульованої провідності кристалів CsPbBr ₃ та CsPbCl ₃ підданих дії X-опромінення. <i>Богдан Павлик, Роман Лис, Роман Гамерник, Микола Гойдало</i> (10)	145

CONTENTS

INFORMATION SYSTEMS AND TECHNOLOGIES

Application of Artificial Neural Network and Fuzzy Inference in the Houseplant Watering Intelligent System. <i>Roman Korostenskyi, Igor Olenych (8)</i>	5
Properties of networks built on digital images. <i>Oleh Kushnir, Oleksiy Kravchuk, Volodymyr Franiv (16)</i>	13
Enhancement of Sensor Panel Tactile Touch Interface. <i>Oleksandr Karpin, Zinovii Liubun, Vasyl Mandziy, Oleh Tereshchuk, Nestor Hotsiy (8)</i>	29
Determining of the best optimizer for a neuronetwork in the development of automatic image tagging system. <i>Andrian Kozynets (10)</i>	37
Research on the efficiency of processing parallel requests by Amazon Web Services using containers. <i>Oleg Sigunov, Lidiia Demkiv (10)</i>	47
Deep Learning Objects Detection through Model Optimization to Accelerate the Process on General-purpose Single-board Platforms. <i>Dmytro Myroniuk, Bohdan Blagitko (12)</i>	57
Problems of using neural networks to predict the price of virtual assets. <i>Andrii Tsemko, Maksym Matskiv (10)</i>	69
Energy conservation as one of the components of the management system for the smart sustainable workspaces. <i>Maksym Yakubovych, Vasyl Lyashkevych, Roman Shuvar (16)</i>	79

MODELING OF PROCESSES AND EFFECTS

The Inverse Gaussian Plume Method for Estimating the Level of Air Pollution. <i>Volodymyr Hura, Lyubomyr Monastyrskiy (16)</i>	95
Method for estimating the Doppler factor of a broadband signal under conditions of multiplicative interference with unknown probability distribution. <i>Pavlo Kostenko, Mykyta Alonkyn, Valeriy Slobodyanuk, Mykhailo Rovetskii, Ivan Rovetskii (12)</i>	111
Computerized Optical Experiments with Portable Spectrometers and Amazon Web Services Cloud Integration. <i>Andriy Krupych, Ivan Karbovnyk (12)</i>	123

MATERIALS FOR ELECTRONIC ENGINEERING

Photosensitive Structures Based on the Silver-doped Porous Silicon and Reduced Graphene Oxide. <i>Igor Olenych, Lyubomyr Monastyrskii, Petro Parandiy, Roman Serkiz, Ivan Karbovnyk (10)</i>	135
Changes in thermally stimulated conductivity of CsPbBr ₃ and CsPbCl ₃ crystals exposed to X-ray irradiation. <i>Bohdan Pavlyk, Roman Lys, Roman Gamernyk, Mykola Goydalo (10)</i>	145

UDC 004.932

APPLICATION OF ARTIFICIAL NEURON NETWORK AND FUZZY INFERENCE IN THE HOUSEPLANT WATERING INTELLIGENT SYSTEM

Roman Korostenskyi , **Igor Olenych** 

*Radioelectronic and Computer Systems Department,
Ivan Franko National University of Lviv,
50 Dragomanov Street, 79005 Lviv, Ukraine*

Korostenskyi, R. & Olenych, I. (2025). Application of artificial neural network and fuzzy inference in the houseplant watering intelligent system. Electronics and Information Technologies, 29, 5-12.
<https://doi.org/10.30970/eli.29.1>

ABSTRACT

Background. The automation of modern technologies and the widespread adoption of smart Internet of Things (IoT) devices not only provide an opportunity to optimize routine processes but also offer innovative solutions within intelligent smart home systems, enhancing the daily comfort of residents. In particular, the integration of artificial intelligence (AI) technologies can expand the functionality of embedded smart home systems and improve their efficiency. Consequently, the development of hardware and software tools for such systems represents a pertinent area of research in the field of smart solutions.

Materials and Methods. The automatic indoor plant irrigation system is implemented using the Raspberry Pi 4 microcomputer and the SEN0308 soil moisture sensor. Control of the irrigation system is achieved through an artificial neural network with two hidden layers and fuzzy logic inference methods. These methods determine the required water volume for irrigation based on soil relative moisture and air temperature.

Results and Discussion. The system operates by periodically analyzing the soil's relative humidity and indoor air temperature, making decisions regarding the activation of the irrigation mechanism based on the collected data. It has been established that implementing artificial intelligence algorithms and fuzzy logic enables automatic regulation of water usage according to environmental conditions, preventing excessive soil moisture. The proposed approach ensures an optimal soil moisture level of 50–60%, suitable for plant growth and development, while also reducing the average monthly water consumption by 25–27%.

Conclusion. The results indicate the high efficiency of using artificial intelligence and fuzzy logic methods for the rational utilization of water resources. Moreover, the proposed algorithms can be implemented on low-power computational platforms commonly used in automation and IoT systems. The proposed intelligent irrigation system demonstrates the potential for functional expansion and the integration of innovative technologies for indoor plant care.

Keywords: Smart home, IoT, embedded systems, irrigation system, artificial neural network, fuzzy logic.

INTRODUCTION

The technological revolution of recent decades has profoundly transformed daily life. Among the key trends driving this progress is the automation of modern technologies, which not only optimizes routine processes but also, through the widespread adoption of the Internet, offers innovative solutions for everyday comfort in intelligent smart home systems [1-3].



© 2025 Roman Korostenskyi & Igor Olenych. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Examining contemporary models of embedded smart home systems reveals key technological components, including centralized control systems and smart sensors. Popular systems such as Google Home, Amazon Alexa, and Apple HomeKit utilize voice control to activate and manage various home devices [4-6]. Integrated solutions encompass intelligent automation systems that regulate lighting, heating, ventilation, and air conditioning. Notably, smart home systems from Crestron or Control4 allow users to schedule and create scenarios for different situations [7,8]. Smart thermostats can automatically adjust home temperatures based on residents' schedules and habits. Continuous monitoring of all smart home systems optimizes energy consumption and enhances energy efficiency. Security systems frequently incorporate video surveillance, motion and door sensors, and technical tools for identification and access control. For instance, Google's Nest Secure provides an integrated system of video surveillance and security sensors with remote control capabilities via smartphone [9]. Furthermore, embedded smart home systems employing unified platforms for integration and management can interact seamlessly with various household devices under the IoT paradigm [10,11].

Most IoT household devices in smart homes are designed to perform single functions, such as watering plants at a predetermined time or raising blinds at sunrise. However, the rapid advancement in microcomputer technology, edge computing, and the availability of tools like TensorFlow Lite, which enables neural network deployment on mobile phones, tablets, televisions, and other embedded systems, opens new possibilities for the development of intelligent smart home systems [12].

This study aimed to explore the integration of artificial intelligence (AI) and fuzzy logic technologies in automatic indoor plant irrigation systems to ensure the efficient use of natural resources. Special attention was given to investigating the potential of low-power computational technologies and artificial neural networks in advancing innovative smart home systems.

MATERIALS AND METHODS

To investigate the efficiency of using artificial intelligence and fuzzy logic technologies in smart home systems, an automatic indoor plant irrigation system was developed based on a Raspberry Pi 4 microcomputer [13] and a SEN0308 soil moisture sensor [14]. The schematic diagram of the irrigation system is shown in Fig. 1.

The system operates by periodically analyzing the relative humidity (RH) of the soil and air, as well as the air temperature, every 8 hours. Based on the collected data, decisions are made regarding the activation of the irrigation mechanism. Since the experimental studies were conducted using a ficus plant, the proposed irrigation system aimed to maintain optimal conditions for this plant. Specifically, in the simplest implementation, automatic watering with 150 ml of water occurred when the soil moisture at a depth of 6–7 centimeters fell below 50% (Fig. 2).

The application of artificial intelligence technology and fuzzy logic algorithms not only determines the necessary volume of water for irrigation but also adapts it to changing environmental conditions, maintaining optimal soil moisture for the ficus within the range of 50–60%. In the initial implementation, the plant irrigation system was controlled using a neural network with two hidden layers, trained on historical data of soil moisture and water usage over 100 epochs. Mean Squared Error (MSE) was employed as the loss function, and the weights of the neural network were optimized using the backpropagation algorithm with an adaptive gradient optimizer [15,16]. Model validation was performed through cross-validation using independent datasets.

Another approach involved using a fuzzy logic inference system with two inputs and one output. This system determined the required water volume for irrigation based on the soil relative moisture and indoor air temperature. The input linguistic variables were "*Soil moisture*" = {"Low" (L), "Moderate" (M), "High" (H)} and "*Indoor temperature*" = {"Cold" (C),

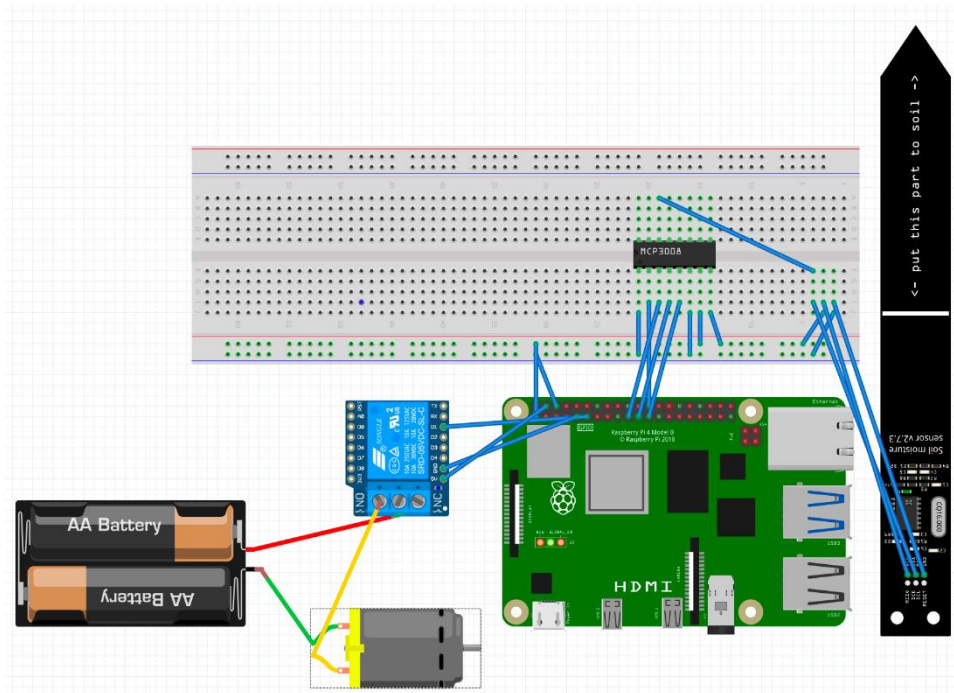


Fig 1. Schematic diagram of the automated indoor plant irrigation system.

"Warm" (W), "Hot" (H)}. The terms of the output linguistic variable "Used water volume" characterized the need for very low (LL), low (L), moderate (M), high (H), and very high (HH) levels of soil hydration. The membership functions of the fuzzy sets corresponding to the terms of these linguistic variables are shown in Fig 3.

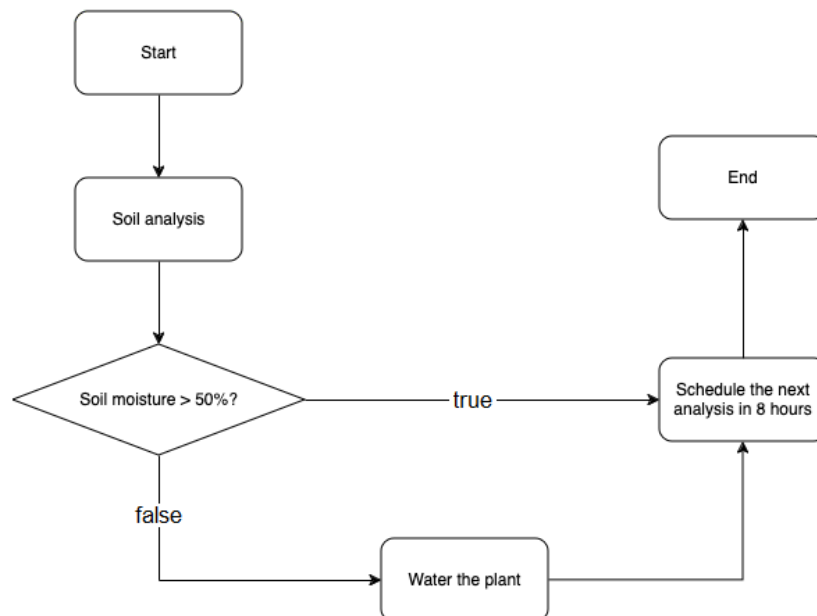


Fig 2. Algorithm of automated irrigation system operation.

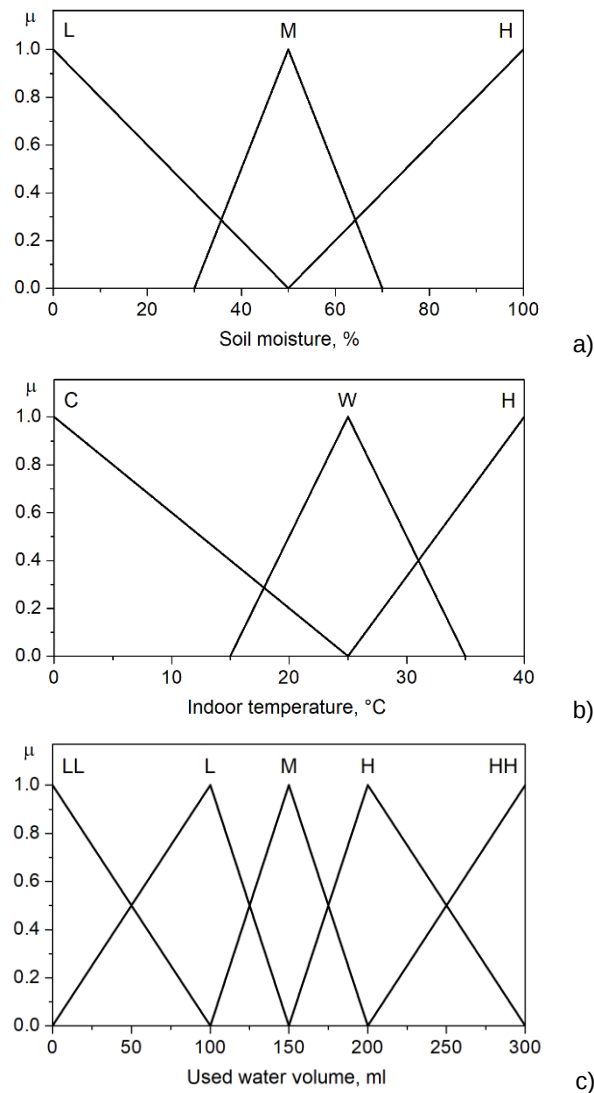


Fig. 3. Membership functions of fuzzy sets characterizing the input linguistic variables "Soil moisture" (a) and "Indoor temperature" (b) as well as output linguistic variable "Used water volume" (c)

The fuzzy rule base was formed in the form IF (sub-condition 1) AND (sub-condition 2) THAT (action). Sub-condition 1 establishes the affiliation of the current value of the soil relative moisture to fuzzy sets characterizing the terms of the linguistic variable "Soil *mois-ture*". Sub-condition 2 is associated with the affiliation of the current value of temperature to terms of the linguistic variable "Indoor *temperature*". The fuzzy sets of the corresponding terms of the output linguistic variable "Used *water volume*" are the rule conclusions.

The base of fuzzy rules for the houseplant watering system is given in Table 1. In general, the fuzzy rules accumulate information from various sources, including experimental data, simulation results, and expert opinions.

The fuzzy logic-based irrigation system was implemented using the Mamdani algorithm, whose main stages include the fuzzification of input variable values, min-activation and max-accumulation of the conclusions from fuzzy production rules, and the defuzzification of the output variable using the centroid method [17,18].

Table 1. The base of fuzzy rules of the houseplant watering system

Soil Moisture	Indoor Temperature		
	L	M	H
L	M	H	HH
M	L	M	H
H	LL	L	M

RESULTS AND DISCUSSION

The primary objective of the developed system is to ensure efficient irrigation of indoor plants while minimizing water consumption and preventing potential harm to the plants. Considering that environmental conditions such as temperature and humidity vary throughout the year, the effectiveness of irrigation can also fluctuate. Therefore, experimental studies were conducted in two stages under different indoor climatic conditions. In the first stage, the relative humidity and air temperature were approximately 60% and 20°C, respectively. During the second stage, the plant was subjected to warmer and drier conditions – with a relative humidity of about 45% and an air temperature of 25°C.

An important parameter for the functioning of the irrigation system is the soil moisture at the time of each watering, as it is essential to maintain the optimal soil moisture level for the plant – approximately 55% for ficus. The temporal variations of soil moisture under different indoor climatic conditions and various irrigation methods are illustrated in Fig. 4.

Integrating artificial intelligence technologies into the irrigation system enables automatic regulation of the water volume used based on environmental conditions, preventing excessive soil moisture. As a result, this approach not only provides optimal conditions for plant growth and development but also ensures rational water consumption (see Table 2). Specifically, a reduction of 25% in average monthly water usage was observed during the first stage of the studies, and a 27% water savings during the second stage.

The intelligent indoor plant irrigation system based on fuzzy logic inference demonstrates similar water usage and soil moisture levels to those of the neural network system (see Fig. 4). The results of testing the irrigation system with fuzzy inference are illustrated in Fig. 5. Based on the analysis of the input–output surface of the fuzzy controller, it was established that the volume of water used for irrigation depends on both soil moisture and air temperature. Specifically, the irrigation system requires a significantly larger volume of water when soil moisture is low and indoor air temperature is high.

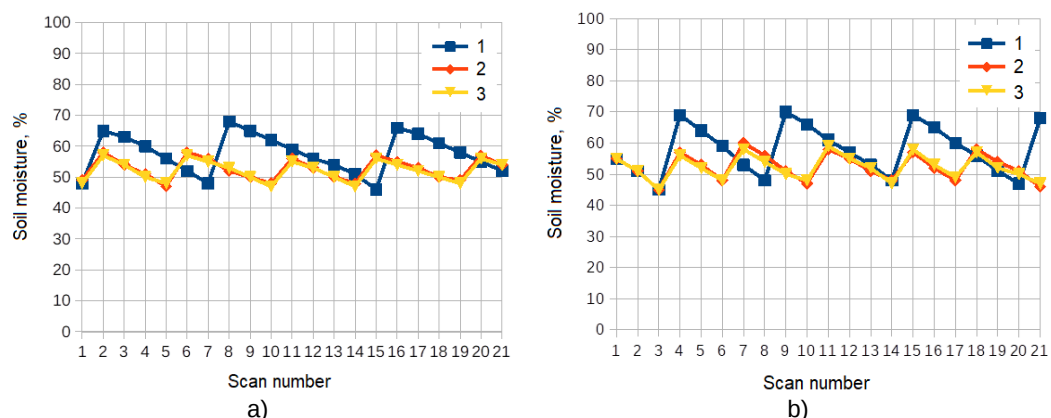


Fig. 4. Time dependence of soil moisture during the first (a) and second (b) stages of the study using the automated irrigation system (1) and one with an integrated artificial neural network (2) and fuzzy inference (3)

Table 2. Average monthly water volume used for irrigating an indoor plant (ficus) under different indoor climatic conditions

Irrigation system type	Used water volume, ml	
	Stage 1 RH=60%, $t=20^{\circ}\text{C}$	Stage 2 RH=45%, $t=25^{\circ}\text{C}$
Automated	2228	3085
Integrated AI	1688	2260
Fuzzy inference	1663	2243

Unlike more complex machine learning algorithms, fuzzy logic methods can be effectively implemented on low-power platforms such as the Raspberry Pi Pico or Arduino, which are widely used in automation and IoT systems. The low cost and low energy consumption of microcontrollers make them an ideal choice for use in budget-friendly and energy-efficient irrigation systems. Moreover, fuzzy logic algorithms allow for the consideration of a wide range of input parameters and enable decision-making that mimics human reasoning.

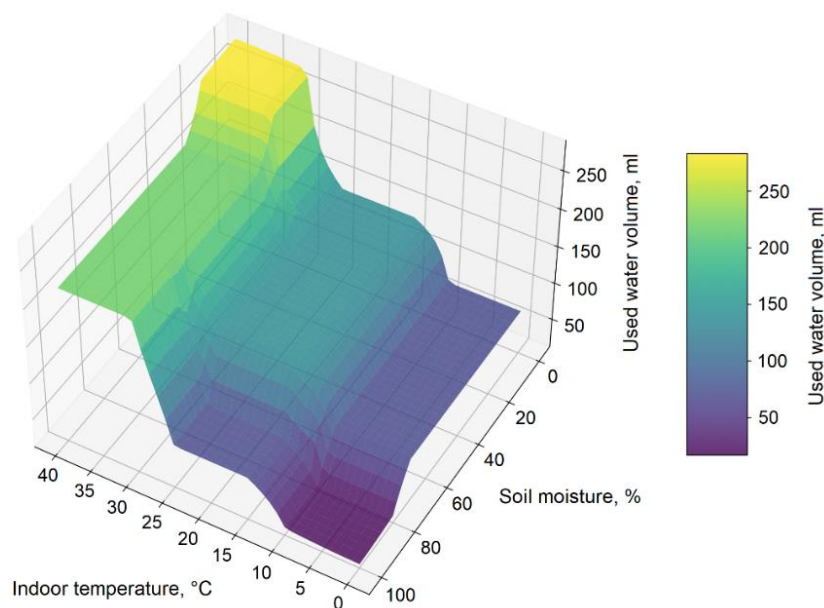


Fig 5. Dependence of used water volume on soil moisture and indoor temperature

CONCLUSION

In this work, an automated indoor plant irrigation system was developed using IoT technologies. The proposed irrigation system enables periodic monitoring of soil moisture and waters the plant as needed. Soft computing methods were employed to determine the necessary amount of water under varying air temperatures and relative soil moisture levels. It was established that integrating an artificial neural network and fuzzy logic inference into the irrigation system allows for maintaining an optimal moisture level in the range of 50–60% and reducing water consumption by 25–27%. An additional advantage of using fuzzy logic inference algorithms is their ability to be implemented on low-power computing

devices, aligning with the IoT paradigm. By integrating new sensors, the proposed intelligent irrigation system has the potential to expand its functionality and implement innovative technologies for indoor plant care.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, [R.K., I.O.]; methodology, [R.K., I.O.]; validation, [R.K.]; investigation, [R.K.]; writing – original draft preparation, [R.K.]; writing – review and editing, [I.O.]; visualization, [R.K., I.O.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Ming, C., Kadry, S., & Dasel, A. (2020). Automating smart Internet of Things devices in modern homes using context-based fuzzy logic. *Computational Intelligence*, 40, e12370. <https://doi.org/10.1111/coin.12370>
- [2] Chakraborty, A., Islam, M., Shahriyar, F., Islam, S., Zaman, H. U., & Hasan, M. (2023). Smart Home System: A Comprehensive Review. *Journal of Electrical and Computer Engineering*, 2023, 7616683. <https://doi.org/10.1155/2023/7616683>
- [3] Robles, R. J., & Kim, T.-H. (2010). Applications, systems and methods in smart home technology: A review. *International Journal of Advanced Science and Technology*, 15, 37–47. <https://api.semanticscholar.org/CorpusID:107641231>
- [4] About Google Home. (2024). <https://home.google.com/about-google-home/>
- [5] Alexa Developer Documentation. (2024). <https://developer.amazon.com/en-US/docs/alexa/documentation-home.html>
- [6] Apple HomeKit. *Developing apps and accessories for the home*. (2024). <https://developer.apple.com/apple-home/>
- [7] Crestron Home. (2024). <https://www.crestron.com/Products/Market-Solutions/Crestron-Home>
- [8] Control4 Whole Home. (2024). <https://www.control4.com/solutions/whole-home>
- [9] Nest Secure User's Guide. (2017). <https://nest.com/support/images/secure-october-2017/Nest-Secure.pdf>
- [10] The Architecture of SmartThings. (2024). <https://developer.smarthings.com/docs/getting-started/architecture-of-smarthings>
- [11] Taiwo, O., & Ezugwu A. E. (2021). Internet of Things-Based Intelligent Smart Home Control System. *Security and Communication Networks*, 2021, 9928254. <https://doi.org/10.1155/2021/9928254>
- [12] Tang, C., Qiu, C., & Xiao, X. (2023). Machine-learning and water energy harvesting based wireless water consumption sensing system in buildings. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 5, 100260. <https://doi.org/10.1016/j.prime.2023.100260>
- [13] Raspberry Pi 4 Model B Datasheet. (2024). <https://datasheets.raspberrypi.com/rpi4/raspberry-pi-4-datasheet.pdf>
- [14] SKU_SEN0308 Specifications. (2023). https://wiki.dfrobot.com/Waterproof_Capacitive_Soil_Moisture_Sensor_SKU_SEN0308
- [15] Adam. (2024). https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam
- [16] Olagoke, I. (2023). Optimizing TFLite Models for On-Edge Machine Learning for Efficiency: A Comparison of Quantization Techniques. *About Medium*. <https://ibrahimgoke.medium.com/optimizing-tflite-models-for-on-edge-machine-learning-for-efficiency-a-comparison-of-quantization-2c0123959cb6>

- [17] Bai, Y., & Wang, D. (2006). *Fundamentals of fuzzy logic control - fuzzy sets, fuzzy rules and defuzzification. Advanced Fuzzy Logic Technologies in Industrial Applications*. Springer.
- [18] Olenych, I. (2022). Smart home climate control system based on fuzzy logic controller. *Electronics and Information Technologies*, 17, 26–35. <https://doi.org/10.30970/eli.17.3>

ЗАСТОСУВАННЯ ШТУЧНОЇ НЕЙРОМЕРЕЖІ ТА НЕЧІТКОГО ВИСНОВКУ В ІНТЕЛЕКТУАЛЬНІЙ СИСТЕМІ ПОЛИВУ КІМНАТНИХ РОСЛИН

Роман Коростенський, Ігор Оленич

*Кафедра радіоелектронних і комп'ютерних систем,
Львівський національний університет імені Івана Франка,
вул. Драгоманова 50, 79005 м. Львів, Україна*

АНОТАЦІЯ

Вступ. Автоматизація сучасних технологій і широке впровадження розумних пристроїв Інтернету речей дає змогу не тільки оптимізувати рутинні процеси, але й запропонувати інноваційні рішення в інтелектуальних системах розумного будинку, які забезпечують повсякденний комфорт мешканців. Зокрема, використання засобів штучного інтелекту може розширити функціональність вбудованих систем розумного будинку і підвищити їх ефективність. Тому розробка програмно-апаратних засобів для таких систем є одним із актуальних напрямків досліджень у галузі розумних рішень.

Матеріали та методи. Система автоматичного поливу кімнатних рослин реалізована на основі мікрокомп'ютера Raspberry Pi 4 та сенсора вологості ґрунту SEN0308. Для керування іригаційною системою було використано штучну нейромережу з двома прихованими шарами та методи нечіткого логічного висновку, які на основі значень відносної вологості ґрунту та температури повітря визначають об'єм води для поливу.

Результати. Функціонування системи полягає у періодичному аналізі відносної вологості ґрунту та температури повітря у приміщенні і прийняті рішення про активацію механізму поливу на основі отриманих даних. Встановлено, що реалізація алгоритмів штучного інтелекту та нечіткої логіки дає змогу автоматично регулювати об'єм використаної води залежно від умов навколишнього середовища та уникати надмірного зволоження ґрунту. Запропонований підхід забезпечує не тільки оптимальний для росту та розвитку рослин рівень зволоження ґрунту у діапазоні 50–60%, але й зменшення середньомісячного об'єму використаної води на 25–27%.

Висновки. Одержані результати свідчать про високу ефективність використання методів штучного інтелекту та нечіткої логіки для раціонального використання водних ресурсів. Крім того, запропоновані алгоритми можуть бути реалізовані на малопотужних обчислювальних платформах, які широко використовуються в системах автоматизації та Інтернету речей (IoT). Запропонована система інтелектуального поливу має потенціал для розширення функціональності та впровадження інноваційних технологій для догляду за кімнатними рослинами.

Ключові слова: Розумний будинок, інтернет речей, вбудовані системи, іригаційна система, штучна нейромережа, нечітка логіка.

Received / Одержано
14 January, 2025

Revised / Доопрацьовано
12 February, 2025

Accepted / Прийнято
15 February, 2025

Published / Оpubліковано
29 March, 2025

УДК 004.6, 004.9, 538.9, 004.932

ВЛАСТИВОСТІ МЕРЕЖ, ПОБУДОВАНИХ НА ЦИФРОВИХ ЗОБРАЖЕННЯХ

Олег Кушнір*, Олексій Кравчук, Володимир Франів

Кафедра оптоелектроніки та інформаційних технологій,
Львівський національний університет імені Івана Франка
вул. Ген. Тарнавського, 107, 79017 м. Львів, Україна

Кушнір, О. С., Кравчук, О. М., Франів, В. А. (2025). Властивості мереж, побудованих на цифрових зображеннях. *Електроніка та інформаційні технології*, 29, 13–28.
<https://doi.org/10.30970/eli.29.2>

АНОТАЦІЯ

Вступ. Відомо, що деякі важливі властивості текстів, написаних природними мовами, можна виявити при відображенні тексту на мережу та вивченні властивостей цієї мережі. Тут слова тексту розглядають як вузли, а їхнє спільне випадання в тексті – як зв'язки.

Матеріали та методи. Проводячи аналогію між текстами та цифровими зображеннями, у цій роботі ми експериментально вивчили та феноменологічно проаналізували властивості мережі, побудованої на зображенні. У цій мережі вузлами є значеннями пікселів, а зв'язки приписують тим значенням пікселів, які просторово прилягають одне до одного на зображенні. При цьому сусідство пікселів визначають радіусом r ($r = 1, 2, \dots$). Ми побудували мережі для інформативного тестового зображення та відповідних зображень, одержаних із початкового зображення шляхом додавання каліброваних порцій білого Гауссового шуму. Проаналізовано і зважені, і незважені мережі, а також порівняно низку практичних методів побудови мережі зображення і додавання шуму до зображення.

Результати та обговорення. Знайдено основні параметри мережі, такі як середній коефіцієнт кластеризації та середня довжина найкоротшого шляху, як функції параметра відносного шуму. Проаналізовано основні якісні та кількісні ознаки цих залежностей. Характеристики мереж критично порівняно з характеристиками лексичних мереж, побудованих на текстах природними мовами (ТПМ) і рандомізованих текстах, одержаних шляхом випадкових перестановок слів у початковому ТПМ. Розглянуто основні подібності та відмінності лексичних мереж і мереж для зображень.

Висновки. Показано, що за середнім коефіцієнтом кластеризації та довжиною шляху мережі, побудовані на основі початкового інформативного зображення та відповідних зашумлених зображень, є «тісними світами». Незважаючи на цю подібність, згадані мережі виявляють низку відмінних особливостей, завдяки яким залежності параметрів мережі зображень від рівня шуму можна використати для розрізнення інформативних і зашумлених зображень. Нарешті, ми проаналізували низку віддалених наслідків наших емпіричних результатів і деякі дані, відомі з літератури. Оскільки і інформативні, і зашумлені зображення виявляють ефект «тісного світу», цей ефект навряд чи доцільно було би пов'язувати з інформаційним навантаженням зображення.

Ключові слова: складні мережі, аналіз і класифікація зображень, розпізнавання зображень, виявлення інформації та шуму, семантика, рандомні моделі.



1. ВСТУП І ПОСТАНОВКА ПРОБЛЕМИ

Технології цифрового опрацювання зображень, комп'ютерного зору та розпізнавання образів, а також опрацювання природної мови [1, 2] є важливими напрямками сучасної інформатики. Вони лежать в основі інноваційних рішень, що впливають на різні сфери нашого життя. Слід зазначити, що алгоритми, підходи та методи, які використовують галузі опрацювання зображень і мови, переважно різні, а випадки впливу напрацювань в одній із галузей на розвиток іншої галузі поодинокі (див. [3, 4]). Це зрозуміло хоча би з того, що тексти є одновимірними, а зображення – двовимірними об'єктами.

Цікавою в цьому плані була спроба [5] розвинути аналогії між зображеннями і текстами та адаптувати методи опрацювання природної мови до аналізу цифрових зображень. При цьому зображення пропонувалося зводити до одновимірного піксельного ряду (який є аналогом часового ряду в класичній обробці сигналів) шляхом його пострічкової або постовпцевої розгортки, а «словами» у відповідному синтетичному «тексті» були числа, що відповідають значенням яскравості пікселів. Основна ідея цього підходу полягала в тому, що математичні прийоми та методики, придатні до аналізу одновимірних об'єктів, простіші та ліпше розвинуті. Як приклад, автори [5] застосовували методи статистичної лінгвістики для вивчення т. зв. «спалахів» (в інших термінах, кластеризації або intermittency) «часового ряду» яскравостей пікселів у «тексті». Було показано, що відповідні кількісні характеристики простого інформативного зображення і шуму помітно різні.

Слід зазначити, що розв'язання задачі розрізнення інформативних зображень від зашумлених (тобто інформативних зображень із внесеними шумами) має вагомі теоретичні та практичні виміри (див., наприклад, огляд [6]). Уявімо стандартну схему, за якою ми виходимо з деякого «безшумного» зображення, що містить всю потрібну нам інформацію. Далі на це зображення багаторкратно та дозовано накладають деякий шум (білий гаусовий шум, імпульсний шум «сіль і перець», мультиплікативний шум тощо), щоразу слідкуючи за деяким кількісним параметром зображення. Хоча зі зростанням частки шуму на перетворених зображеннях все ще можна буде ідентифікувати важливі для нас об'єкти або структури, загалом ці зображення міститимуть все менше інформації. Подальше зростання рівня шуму значно ускладнить або й зробить неможливим розпізнавання елементів початкового зображення, зрештою повністю приховавши його інформаційну складову. Так от, завдання полягає в тому, аби знайти деякий кількісний параметр, який виявиться помітно різним для зображень із різними частками шуму. Завданням-максимумом є розв'язання оберненої задачі – кількісного знаходження наперед невідомої частки шумів на зображенні, виходячи зі значення згаданого параметра.

Звісно, описану задачу можна розв'язувати за стандартними підходами опрацювання зображень [1], проте всі вони виявляють певні нестрогості та недоліки, так чи інакше пов'язані з проблемами розрізнення шумів на тлі складних і дрібних інформативних деталей зображення (див. [6, 7]). Це стосується і аналізу гістограм зображень або їхньої інформаційної ентропії, і відношення сигнал/шум (ВСШ або signal-to-noise ratio), і інших підходів [1]. Так, хоча математичні та алгоритмічні основи підходу ентропії нормалізованої гістограми бездоганні і це справді автоматизований підхід (див. [1]), але ентропія інваріантна щодо рандомізації зображення, тобто її величина та сама і для початкового (припустимо, інформативного) зображення, і для його повністю рандомізованого аналогу, який не містить жодної інформації. Зі схожих причин для розпізнавання інформативних і шумових зображень не годиться й контраст.

З іншого боку, визначення дисперсії яскравостей і ВСШ не є повністю автоматизованим підходом, бо їхнє оцінювання часто засноване на «ручному» виділенні «ділянок сигналу» і відносно однорідної ділянки (або ділянок) зображення, де

відсутній інформаційний сигнал і всі зміни яскравості викликані виключно шумами. Крім того, в ідеалі ВСШ можна знайти найточніше, порівнюючи зашумлене зображення із зображенням, на якому шуми завідомо відсутні. У дещо інших термінах, ВСШ дає швидше відносну оцінку шуму для двох зображень, але не абсолютну оцінку шуму для єдиного зображення, що є однією з наших задач.

Отже, подальша розробка методів оцінювання рівнів шумів на зображеннях є актуальною. Дане дослідження продовжує розвивати напрям перенесення методів аналізу текстів на зображення [5]. Нижче ми зосередимося на короткому описі методики вивчення мережевих характеристик текстів і на аналізі залежностей мережевих характеристик цифрових зображень від відносного вмісту гаусового шуму на них.

2. СКЛАДНІ МЕРЕЖІ ТА МЕРЕЖЕВІ ПАРАМЕТРИ. ЛІНГВІСТИЧНІ МЕРЕЖІ

Мережі – це універсальний інструмент моделювання складних систем із виразними взаємозв'язками між їхніми елементами. Мережа складається з вузлів, які є деякими об'єктами, та зв'язків, які відображають взаємодії між ними. Перерахуємо головні мережі параметри, які ми будемо аналізувати (точні означення див., наприклад, у [8, 9]):

- кількість V вузлів у мережі;
- середній коефіцієнт кластерності C мережі;
- середня довжина l найкоротшого шляху між вузлами мережі;
- середня довжина l_w найкоротшого шляху для зваженого варіанту мережі;
- середній ступінь k вузлів мережі;
- середня кратність M зв'язків у зваженій мережі.

Тут параметри C , l , l_w , k і M усереднюють локальні параметри C_i , l_i , l_{wi} , k_i і M_i , що стосуються вузлів i ($i = 1..V$). Зазначимо, що за означенням кратність усіх зв'язків у незваженій мережі дорівнює одиниці ($M_{ij} = M = 1$). Нарешті, на противагу до незваженої (бінарної) мережі, найкоротші шляхи між вузлами в зваженій мережі ми модифікуємо так, що довжина $l_{w,ij}$ зв'язку між деякими найближчими сусідніми вузлами i та j дорівнює $l_{w,ij} = 1/M_{ij}$, де M_{ij} – кратність зв'язку між цими вузлами. Іншими словами, чим більшою кількістю зв'язків пов'язані вузли i та j , тим менша «зважена відстань» між ними. Ця дефініція стає зрозумілішою, якщо залучити аналогію. Так, логічно припустити, що «відстань» між студентом та його батьком, між якими є багатократні зв'язки, коротша за відстань між цим студентом і його викладачем, або що вузли деякої мережі постачання, зв'язані трубопроводами більшого діаметру, є «ближчими», аніж у разі тонших трубопроводів (принаймні швидкість або ефективність постачання в першому випадку буде вищою). Водночас, довжина l_{ij} зв'язку між найближчими сусідами в незваженій мережі дорівнює $l_{ij} = 1$,

Одна з важливих властивостей багатьох складних мереж – це ефект «тісного світу». Він описується двома ознаками: малим середнім шляхом у мережі, порівняним зі шляхами, типовими для випадкових графів (ВГ), і набагато вищим коефіцієнтом кластерності, ніж для ВГ. Порівнюючи мережу «тісного світу» та ВГ із однаковими розмірами, матимемо властивості $l \sim l_{ВГ}$ і $C \gg C_{ВГ}$. Крім того, складні мережі часто виявляють важкий (степеневий) хвіст масового розподілу ймовірності $\text{pmf}(k_i)$ ступенів вузлів. Такі мережі називають безмасштабними, що означає відсутність характерного просторового масштабу в мережі.

Лінгвістичні мережі – це частковий випадок мереж, у яких вузлами є лінгвістичні елементи (переважно слова), а зв'язки найчастіше відображають сусідство слів у тексті (в т.ч. появу слів у тому ж реченні), або зрідка – семантичні відносини між ними. Дослідження [9–11] засвідчили, що лінгвістичні мережі також є «тісними світами» і виявляють безмасштабність. В останнє десятиліття лінгвістичні мережі привертати

значну увагу дослідників. Зокрема, їх вивчали з метою розрізнення природних мов, типів текстових документів і якісних характеристик текстів [12–15]. Що важливо, було проаналізовано й проблему розрізнення ТПМ (які є семантично наповненими) від рандомізованих (тобто семантично порожніх) текстів [16–19]. Тоді як ТПМ зі змістом фактично можна вважати деяким аналогом інформативних зображень, рандомізовані тексти представляють собою випадкові послідовності лінгвістичних символів, а тому є аналогами шумових зображень. Хоча автори [16, 18, 19] і констатували відсутність значних успіхів у розрізненні ТПМ і деяких конкретних рандомізованих моделей природної мови за мережевими методами, все ж цікаво було би спробувати застосувати ці підходи до розв'язання задачі розрізнення інформативних і шумових зображень.

3. МЕТОДИКА ДОСЛІДЖЕНЬ

3.1. Побудова мережі на цифровому зображенні

Ми пропонуємо застосувати мережевий підхід до аналізу цифрових зображень, у рамках якого різні числові значення яскравостей пікселів зображення (аналоги слів або word types у тексті) є вузлами мережі, а зв'язки вузлів відповідають їхньому сусідству на зображенні. «Сусідами» деякого пікселя називають усі пікселі на зображенні, що перебувають на відстані («радіусі») r від цього пікселя (наприклад, $r = 1, 2, \dots$). Далі можна використати такі мережеві параметри як коефіцієнт кластеризації, середню довжину шляху та ін. для оцінювання рівня шуму на зображенні та спробувати відрізнити інформативні зображення від шумових.

Найперше слід вирішити стандартну задачу відображення (mapping) цифрового зображення на мережу. Тут існує два можливі виходи.

1. У дусі підходу [5], ми можемо одразу перетворити двовимірний масив рівнів яскравості пікселів (матрицю цифрового зображення – див. Рис. 1) на одновимірний масив («текст»), вважаючи ці яскравості аналогами слів у тексті. Це можна здійснити пострірковою (або постовпцевою) розгорткою зображення в одновимірний масив чисел-«слів». У разі постріркової розгортки матимемо «текст» на зразок «42 14 252 24 14 63 52 42 52 41 67 64 12 ...» (див. Рис. 1). Далі можна застосовувати стандартні означення лінгвістичної мережі (див. [9]) і будувати цю мережу, виходячи із сусідства значень пікселів у сформованому «тексті». Проте такий спосіб побудови мережі виявляє проблеми спотворення сусідства пікселів, які мають два вияви:

а) пропуски сусідів – тобто пікселів, які були сусідніми в зображенні, проте не є такими у створеному нами синтетичному «тексті»; у разі розгортання зображення по рядках прикладами будуть перші пікселі зліва «12» і «15» у рядках №3 і №4 (див. Рис. 1); на додаток, пікселі, які є діагональними сусідами в зображенні, також перестануть бути сусідами в такому «тексті»;

б) поява оманливих сусідів – пікселів, які не були сусідніми в зображенні, проте стали такими в «тексті»; при розгортанні зображення по рядках прикладами будуть останній піксель справа «64» на зображенні в рядку №2 і перший піксель зліва «12» у наступному рядку №3 (див. Рис. 1).

2. Дещо технічно складнішою альтернативою є урахування істинного 8-сусідства на Рис. 1а (або ширшого сусідства на Рис. 1б) пікселів на зображенні [1] (див. Рис. 1а). Червоні квадрати на Рис. 1а і Рис. 1б ілюструють відповідно випадки радіусів сусідства $r = 1$ і $r = 2$. Ми створили програму, яка обліковує та виписує всі сусідства пікселів на зображенні (див. Рис. 1а), а далі записує їх до «тексту», формат якого дещо інший: «42 14. 42 252. 42 52.». Цей «текст» служить виключно для того, аби врахувати сусідства значень пікселів; сусідами ж у ньому ми вважатимемо лише ті пікселі, які трапляються в тому ж реченні (тобто в наведеному прикладі зв'язаними будуть вузли «42» і «13», «42» і «252» тощо). Як і в методі (1), цей формат

42	14	252	24	14	63
52	42	52	41	67	64
12	35	146	53	54	53
15	37	41	42	211	11
63	42	75	13	67	123
13	13	52	53	31	53

(а)

42	14	252	24	14	63
52	42	52	41	67	64
12	35	146	53	54	53
15	37	41	42	211	11
63	42	75	13	67	123
13	13	52	53	31	53

(б)

Рис. 1. Схематичне представлення зображення: на панелях (а) і (б) червоними квадратами обмежено всіх сусідів пікселів з яскравостями «42» і «146», які виділені зеленим фоном.

Fig. 1. Schematic representation of an image: in panels (a) and (b) all neighbors of the pixels with brightnesses of “42” and “146”, which are highlighted by a green background, are bounded by red squares.

«тексту» дає змогу використовувати існуючі програми для розрахунку мережевих характеристик звичайних текстів.

Отже, метод (2) визначення сусідства усуває всі проблеми спотворення сусідства пікселів за методом (1), описані вище. Звісно, масштаби цих спотворень не мали би бути значними, проте впливу факторів (а) і (б) слід або уникнути, або оцінити його – прямо – або опосередковано, порівнюючи результати для мереж, створених за методами (1) і (2).

Кількість вузлів у мережі, побудованій на зображенні, – це повний набір різних яскравостей пікселів («словник»), що трапляються в ньому. Так, у 8-бітному зображенні може бути максимум $V = 256$ вузлів. Нарешті, зазначимо, що вплив радіуса r на експериментальні дані для мережевих параметрів є окремим предметом інтересу.

3.2. Накладання шуму на зображення

Для дослідження було обрано стандартне 8-бітне тестове RGB-зображення «Lena» розміром 512x512 пікселів (див. Рис. 2а), яке аналізували по червоному каналу. Природно вважати це початкове зображення вільним від шумів або принаймні таким, шумами в якому можна знехтувати, тобто вважати, що стандартне відхилення яскравостей пікселів у ньому зумовлене тільки інформативною складовою.



Рис. 2. Початкове досліджуване зображення (а) і приклади зображень із ґаусовим шумом, додатково накладеним на зображення на панелі (а): робочий параметр шуму *noise* (див. пояснення в тексті) складає $3,86 \cdot 10^{-3}$ (б), 0,974 (в) і 60,3 (г).

Fig. 2. Original test image (a) and examples of images with Gaussian noise additionally superimposed on the image shown in panel (a): the operating parameter *noise* (see explanations in the text) is equal to $3.86 \cdot 10^{-3}$ (б), 0,974 (в) і 60,3 (г).

Далі на початкове зображення дозованими порціями накладали білий ґаусовий шум: до яскравості кожного пікселя на зображенні додавали нормально розподілену випадкову величину з нульовим середнім значенням і різними стандартними відхиленнями σ як аргументами функції шуму `numpy.random.normal(0,  $\sigma$ , img.shape)` у Python, де змінна `img.shape` стосується розмірності зображення (див. ілюстрації на Рис. 2б–г). Для зручності величину ґаусового шуму *noise* виражали у відносних одиницях ($noise = \sigma/\sigma_0$) – як частку стандартного відхилення σ_0 яскравостей пікселів початкового зображення, тобто аргумент σ функції шуму визначали за формулою

$$\sigma = \sigma_0 noise,$$

де для зображення «Lena» маємо $\sigma_0 \approx 60,3$.

Робочу змінну *noise* змінювали в діапазоні 0..60, верхньої межі цього діапазону ($noise \approx \sigma_0$ і $\sigma \approx \sigma_0^2$) цілком вистачало для того, аби шум у зображенні домінував, а інформативна складова зображення була повністю пригнічена (див. Рис. 2г). Для докладного вивчення впливу шуму на мережеві параметри було згенеровано 60 зашумлених зображень з різними параметрами *noise* білого шуму.

Зазначимо, що ми розглядали два альтернативні алгоритми накладання шуму:

А. Режим «обмеження» значень пікселів: якщо після додавання шуму деякі з цих значень у матриці 8-бітного зображення стають від'ємними, то їм присвоюють значення «0»; якщо ж значення стають більшими за «255», то їх прирівнюють до «255».

Б. Режим «відсутності обмеження» значень пікселів: якщо після додавання шуму деякі значення пікселів виходять за межі [0; 255], то їх переозначають за принципом «періодичних граничних умов» (див., наприклад, [20]); наприклад, неприпустиме значення «256» перетворюють на «0», «257» – на «1» тощо; так само значення «-1» перетворюють на «255», значення «-2» – на «254» і т.д.

Хоча інтуїтивні міркування підказують, що режим обмеження (А) може дещо порушити випадковий характер процедури зашумлення зображення, ми все ж аналізували цей режим і потім порівнювали відповідні дані з даними для альтернативного режиму (Б).

4. РЕЗУЛЬТАТИ ТА ЇХНІЙ АНАЛІЗ

На Рис. 3–5 представлено основні результати для мережевих властивостей, одержані за умови радіуса сусідства $r = 1$, а Таблиця 1 підсумовує кількісно основні мережеві дані для наших зображень. Логарифмічну шкалу по осях абсцис на Рис. 3–5 використано для зручності одночасного спостереження за ділянками рівнів шумів, які істотно відрізняються. Відповідно, точку $noise = 0$ для можливості відображення на графіках замінено на $noise = 10^{-5}$. Зокрема, на Рис. 3 показано залежності мережевих параметрів від відносної кількості ґаусового шуму на зображенні для випадку, коли сусідство пікселів означають за методом (1) і не обмежують значення пікселів при накладанні шуму (метод (Б)). Тут зображення розгортати по стовпцях. Зазначимо, що у разі здійснення розгортки по рядках результати для мережевих параметрів схожі до даних Рис. 3. На Рис. 4 і Рис. 5 представлено ті самі залежності для альтернативного випадку, коли сусідство пікселів встановлюють за методом (2). Зокрема, дані Рис. 4 відповідають відсутності обмеження значень пікселів при зашумленні (тобто періодичним граничним умовам), а дані Рис. 5 – наявності такого обмеження.

Якщо трохи спрощено говорити про головні спільні якісні риси залежностей на Рис. 3–5, то всі вони містять початкову ділянку малих шумів, на якій мережевий параметр слабо залежить від рівня шуму, ділянку різких змін (зростання або спадання) цього параметра, а також прикінцеву ділянку практичної незмінності параметра, що відповідає найвищим рівням шумів.

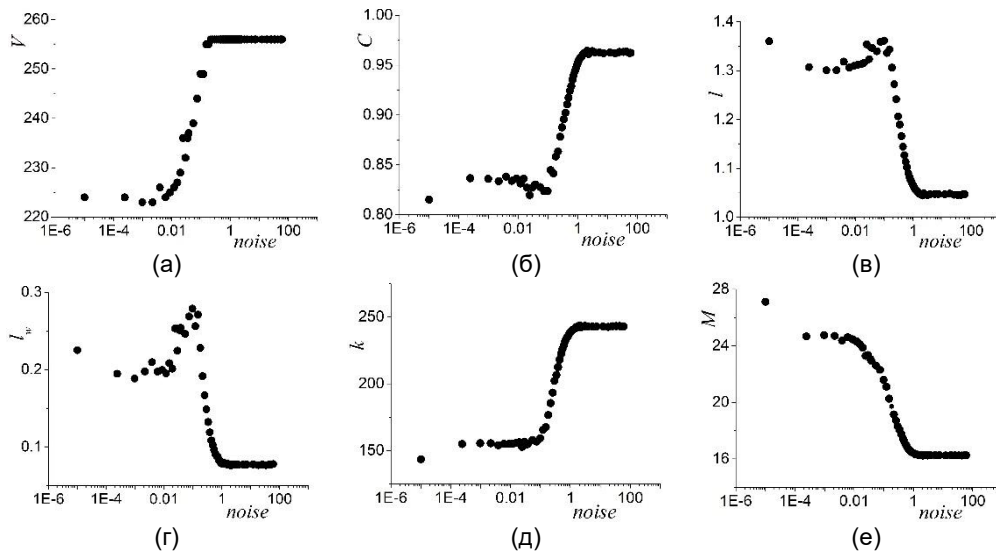


Рис. 3. Залежності основних мережевих параметрів зображення від відносної кількості *noise* гаусового шуму на ньому: означення сусідства яскравостей пікселів за методом (1); значення пікселів при накладанні шуму не обмежують. Тут і на всіх інших рисунках опрацьовують червоний канал початкового зображення, а радіус сусідства $r = 1$,

Fig. 3. Dependences of the main parameters of the image network on the relative amount *noise* of Gaussian noise in the image: definition of the neighborhood of pixel brightnesses is given by the method (1); pixel values are not limited when applying noise. In this and in all the other figures, the red channel of the original image is processed, and the neighborhood radius is $r = 1$,

На початковій ділянці шумів кількість V вузлів мережі $V \approx 223..225$, що є характеристикою конкретного початкового зображення, а на кінцевій ділянці вона зростає до максимально можливої кількості яскравостей у 8-бітному зображенні,

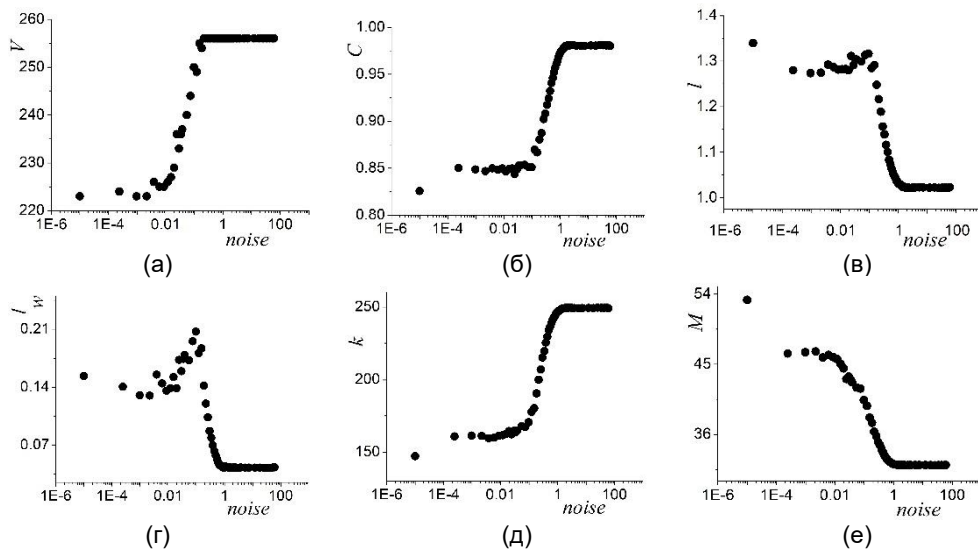


Рис. 4. Залежності основних мережевих параметрів зображення від відносної кількості *noise* гаусового шуму на ньому: означення сусідства яскравостей пікселів за методом (2); значення пікселів при накладанні шуму не обмежують.

Fig. 4. Dependences of the main parameters of the image network on the relative amount *noise* of Gaussian noise in the image: definition of the neighborhood of pixel brightnesses is given by the method (2); pixel values are not limited when applying noise.

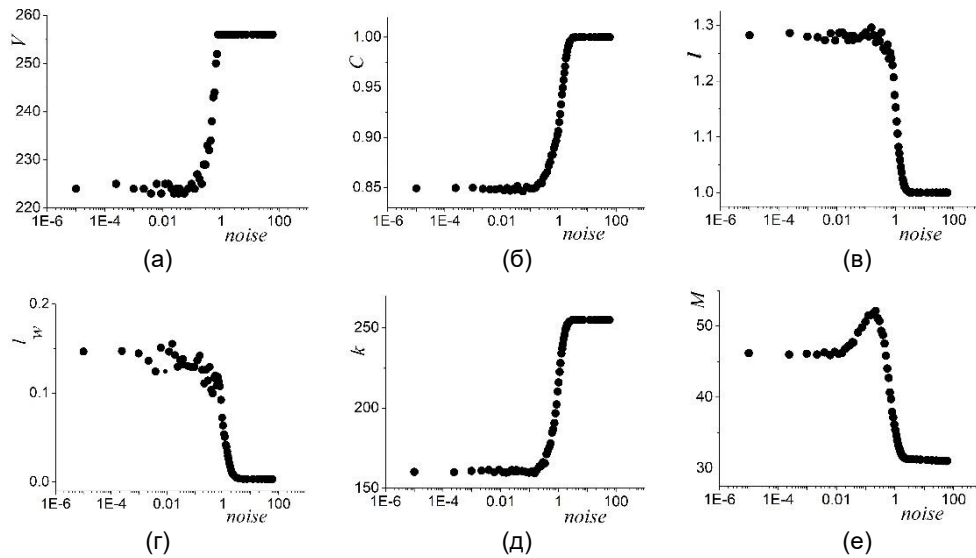


Рис. 5. Залежності основних мережових параметрів зображення від відносної кількості *noise* гаусового шуму на ньому: означення сусідства яскравостей пікселів за методом (2); значення пікселів при накладанні шуму обмежують.

Fig. 5. Dependences of the main parameters of the image network on the relative amount *noise* of Gaussian noise in the image: definition of the neighborhood of pixel brightnesses is given by the method (2); pixel values are limited when applying noise.

тобто 256 (див. Рис. 3а–5а). Звісно, що остання ситуація завдячує домінуючому впливові гаусового шуму, який завдяки випадковому характеру породжує всі ті значення пікселів, яких не було на початковому зображенні. Коефіцієнт кластерності C для безшумових зображень складає $\sim 0,8$, а для максимально зашумлених він сягає значень 0,96 для мереж, побудованих за означенням сусідства (1), і 0,98 або навіть 1,00 для мереж із означенням (2) відповідно у випадках відсутності або наявності обмеження значень пікселів під час зашумлення (див. Рис. 3б–5б і Таблицю 1). Це означає екстремально високу кластеризацію вузлів або транзитивність випадкової мережі (див. [8]): всі сусіди деякого вузла є також сусідами один одного. Зростання C при великих значеннях параметра шуму *noise* є непрямым наслідком випадкового, невибіркового характеру просторового поєднання різних яскравостей у відповідних мережах (див. також нижче). Водночас, у мережі для початкового зображення багато яскравостей «не поєднуються».

Середня довжина найкоротшого шляху l у мережах, побудованих для початкового зображення за всіма методами, дорівнює $\sim 1,3$, а для мереж максимально зашумлених зображень на Рис. 3в–5в вона близька до одиниці (Таблиця 1). Умови $l \approx 1$ або $l = 1$ означають екстремально малі шляхи в мережі, коли майже всі або всі значення пікселів у білому шумі є найближчими сусідами один одного. На відміну від коефіцієнта кластерності C , найкоротший шлях l у мережі зменшується зі зростанням рівня шуму в зображенні. Як і в поясненнях для параметра C , це означає невибіркові зв'язки буквально всіх вузлів-пікселів із довільними яскравостями.

Середні шляхи l_w для зважених мереж істотно менші за відповідні шляхи l для незважених мереж. Це природний наслідок нашого означення шляху в зваженому випадку: чим більше зв'язків існує між деякими вузлами, тим коротший шлях між ними. У кількісному плані маємо $l_w \sim 0,2$ для мереж початкового зображення і $l_w \approx 0,08$; 0,04 і 0,003 відповідно для мереж білого шуму на Рис. 3г–5г (див. Таблицю 1). Останні екстремально низькі значення сигналізують про те, що всі вузли в такій випадковій мережі надзвичайно тісно (багатократно) пов'язані один із одним.

Цікаво, що грубий статистичний аналіз кратності зв'язків між вузлами в повністю випадковій мережі (див. нижче) та відповідна формула

$$l_w \approx 1/M \quad (1)$$

дають оцінку $l \approx 1/32 \approx 0,03$, причому випадковий характер усіх зв'язків і гіпотетична статистична еквівалентність усіх вузлів у мережі мали би означати $l_{w,ij} \approx l_w$ і $M_{ij} \approx M$. Наведена вище оцінка найближча до значення l_w для мережі з означенням сусідства яскравостей пікселів за методом (2) і за відсутності обмеження значень пікселів при накладанні шуму (0,04 – див. Рис. 4г і Таблицю 1). Та ж формула (1) і значення $M = 16$ для кратності у випадку методу побудови мереж (1) і методу зашумлення (Б) (див. Рис. 3а) дають $l_w \approx 0,063$, що близьке до експериментального значення 0,08. Водночас, для випадку методів (2) і (А) маємо значення $l_w \approx 0,003$ (див. Таблицю 1), яке далеке від грубої теоретичної оцінки за формулою (1) і даними експериментів для кратності ($M = 31$). Нарешті, причини спадання довжини шляху l_w для зважених мереж за умови поступового зашумлення зображення подібні до причин для незважених мереж.

При низьких шумах середній ступінь k вузлів всіх трьох типів мереж на Рис. 3д–5д складає ~ 160 , а за умов максимального зашумлення він дорівнює ~ 250 (див. Таблицю 1). Пояснення зростання параметра k для високих рівнів шуму аналогічне до пояснень для коефіцієнта C , а тривіальна якісна оцінка k для повністю випадкової мережі (255 сусідів) найточніше виконується для мережі, побудованої за методом (2) у разі обмеження значень пікселів при зашумленні (Рис. 5д).

Кратність M зв'язків у зважених мережах інформативних зображень на Рис. 3е–5е наближено дорівнює відповідно 25, 47 і 46, а шумових зображень – 16, 32 і 31 (див. Таблицю 1). Зменшення кратності M з підвищенням рівня шуму можна інтерпретувати так. Очевидно, що кожному зображенню та його зваженій мережі відповідає деяка загальна кількість зв'язків n , яка залежить від розмірів зображення ($i = 1..L_1$, $j = 1..L_2$) і радіуса сусідства r . З урахуванням кратності зв'язків матимемо

$$n = \sum_{i,j} (j>i) n_{ij} M_{ij}, \quad (2)$$

де $n_{ij} = 1$ або $n_{ij} = 0$ відповідно якщо вузли i та j зв'язані або не зв'язані;

M_{ij} – це кратність зв'язку вузлів i та j .

Зі зростанням рівня шумів у зображенні все більше вузлів у мережі матимуть зв'язки, тобто все більше чисел n_{ij} у сумі (2) стануть ненульовими. Оскільки повне число n зв'язків при цьому незмінне, то кратності M_{ij} відповідно зменшаться. У граничному випадку повністю зашумленого зображення всі вузли будуть зв'язаними, а за додаткової гіпотетичної умови статистичної еквівалентності всіх зв'язків їхні кратності M_{ij} стануть статистично нерозрізними ($M_{ij} = M$ для будь-яких i та j).

Тепер грубо оцінимо число M для мереж максимально зашумлених зображень. Базовим тут є припущення про повністю випадковий характер цих зображень, тобто просторового розміщення всіх їхніх елементів, а також випадковий характер відповідних мереж і зв'язків у них. Замість строгого розгляду ґаусового розподілу ймовірності появи різних значень пікселів у зображенні (ґаусової нормалізованої гістограми) вважаємо, що цей розподіл рівномірний, тобто маємо гіпотетичну границю нормального розподілу за умови безмежно великого стандартного відхилення. У такому разі всі пікселі, вузли мережі і зв'язки між ними статистично однакові. Зокрема, кожен вузол у такій мережі зв'язаний із кожним іншим вузлом, а кратності цих зв'язків однакові.

Таблиця 1, Параметри мереж, побудованих на досліджуваних зображеннях (радіус сусідства $r = 1$).Table 1, Parameters of the networks built on the images under study (neighborhood radius $r = 1$).

Мережевий параметр	Кількісні дані	Спосіб побудови мережі		
		Означення сусідства (1), метод зашумлення (Б)	Означення сусідства (2), метод зашумлення (Б)	Означення сусідства (2), метод зашумлення (А)
Кількість V вузлів у мережі	V на ділянці низьких шумів	223..225	223..225	223..225
	V на ділянці високих шумів	256	256	256
	Діапазон <i>noise</i> різких змін V	0,01..0,2	0,01..0,2	0,05..0,8
	Діапазон <i>noise</i> піку V	–	–	–
Середній коефіцієнт кластерності C мережі	C на ділянці низьких шумів	0,83..0,85	0,83..0,85	0,83..0,85
	C на ділянці високих шумів	0,96	0,98	1,00
	Діапазон <i>noise</i> різких змін C	0,1..1,5	0,1..1,5	0,15..3,0
	Діапазон <i>noise</i> піку C	–	–	–
Середня довжина l найкоротшого шляху між вузлами мережі	l на ділянці низьких шумів	~ 1,3	~ 1,3	~ 1,3
	l на ділянці високих шумів	1,05	1,02	1,00
	Діапазон <i>noise</i> різких змін l	0,1..2,0	0,1..1,6	0,2..3,0
	Діапазон <i>noise</i> піку l	0,02..0,17	0,02..0,16	–
Середня довжина l_w найкоротшого шляху для зваженої мережі	l_w на ділянці низьких шумів	0,20	0,15	0,14
	l_w на ділянці високих шумів	0,08	0,04	0,003
	Діапазон <i>noise</i> різких змін l_w	0,1..1,0	0,1..1,0	0,12..3,0
	Діапазон <i>noise</i> піку l_w	0,02..0,17	0,02..0,16	–
Середній ступінь k вузлів мережі	k на ділянці низьких шумів	155	160	160
	k на ділянці високих шумів	243	249	255
	Діапазон <i>noise</i> різких змін k	0,08..1,4	0,09..1,4	0,2..2,5
	Діапазон <i>noise</i> піку k	–	–	–
Середня кратність M зв'язків у зваженій мережі	M на ділянці низьких шумів	25	47	46
	M на ділянці високих шумів	16	32	31
	Діапазон <i>noise</i> різких змін M	0,01..1,0	0,01..1,0	0,2..2,0
	Діапазон <i>noise</i> піку M	–	–	0,02..0,5

Скільки разом зв'язків є у зваженій мережі шумового зображення, яке містить $512 \cdot 512 = 2^{18}$ пікселів? Якщо не враховувати крайових ефектів (працювати в грубому наближенні «безмежно великого» зображення), то за умов дефініції сусідства (2) і радіуса сусідства $r = 1$ кожен піксель у зваженій мережі зв'язаний із $8 = 2^3$ сусідніми пікселями. Тому разом буде $2^{18} \cdot 2^3 \approx 2^{21}$ зв'язків, а на кожен із $256 = 2^8$ вузлів мережі 8-бітного зображення припадатиме $2^{21}/2^8 \approx 2^{13}$ зв'язків. Тут не враховано, що зв'язок пікселя з яскравістю i із тим же пікселем i неможливий, проте похибка такого наближення незначна.

2^{13} зв'язків деякого довільного вузла нашої гіпотетичної зваженої мережі статистично рівномірно розподілені між $255 \approx 2^8$ іншими вузлами, а тому кількість усіх зв'язків цього вузла (2^{13}) дорівнює добутку кількості його вузлів-сусідів (2^8) і кратності M його зв'язків із кожним іншим вузлом. Отже, кратність M кожного зв'язку в зваженій мережі такого гіпотетичного шумового зображення наближено дорівнює $2^{13}/2^8 = 2^5 = 32$. Цікаво, що це число, одержане у вкрай грубому наближенні рівномірної гістограми, фактично збігається з даними, одержаними для обох мереж, побудованих на зображенні за методом (2) (див. Рис. 4е і 5е), а параметр M для мережі, побудованої за методом (1) (див. Рис. 3е), є вдвічі меншим.

Зазначимо, що режим обмеження значень пікселів сприяє більш рівномірному «заселенню» екстремально малих і великих значень «0» і «255», порівняно з гаусовим розподілом, тобто розподіл значень пікселів стає дещо ближчим до рівномірного, для якого й було проведено наближені розрахунки. Попри те, дані для кратності зв'язків у цій зваженій мережі істотно відхиляються від результату розрахунку за формулою (1).

Перейдемо до аналізу діапазонів різких змін мережевих параметрів зашумлених зображень, унаслідок яких початкове інформативне зображення перетворюється на шум (див. Таблицю 1). На тлі деяких ознак універсальності, ці діапазони залежать і від самого параметра, і від способу побудови мережі та зашумлення зображення. Дещо спрощуючи ситуацію, бачимо, що всі мережеві параметри, крім кількості вузлів V і кратності M зв'язків у зважених мережах, швидко змінюються в діапазоні $\sim 0,1..1,0$, а діапазони помітних змін параметрів V і кратності M трохи зміщені і/або ширші (відповідно, $\sim 0,01..0,5$ і $\sim 0,01..1,0$).

Крім ділянки різкого зменшення для істотно зашумлених зображень, параметри l , l_w і M виявляють вузькі піки безпосередньо перед цією ділянкою для деяких способів побудови мереж і зашумлення зображень (див. панелі (в), (г) і (е) на Рис. 3–5). Тому залежності цих параметрів від рівня шуму мають більш чи менш виразний немонотонний характер. Діапазони змінної *noise*, у яких спостерігаємо ці піки, показано в Таблиці 1. Причини появи згаданих піків потребують докладнішого вивчення, проте можна помітити, що діапазони піків приблизно відповідають діапазнам, у яких різко змінюється кількість вузлів V . Можливо, існують деякі непрямі механізми впливу змін V на появу згаданих піків.

Чи не найперше практичне питання, пов'язане з оцінюванням наперед невідомого рівня шумів у деякому зображенні, таке: наскільки сильно впливає зашумлення зображення на його мережеві параметри? У Таблиці 2 представлено відповідні результати лише для параметрів C , l і k за умови різних радіусів сусідства $r = 1..5$. Тут фігурують абсолютні ($\Delta C' = C(60) - C(0)$ тощо) і відносні ($\Delta C'' = (C(60) - C(0))/C(0)$, % тощо) відмінності C , l і k , узяті для початкового інформативного (*noise* = 0) та істотно зашумленого (*noise* = 60) зображень. Із даних Таблиці 2 випливає, що задачу розрізнення інформативних і шумових зображень на основі їхніх мережевих властивостей слід вирішувати за умови найменших радіусів сусідства.

У плані порівняння властивостей мереж, побудованих за різними методами і для зображень, зашумлених за різними способами, слід зазначити відносно стійкість мережевих характеристик до цих факторів: хоча при цьому й спостерігаємо деякі кількісні зміни в залежностях мережевих параметрів від рівня шуму, проте всі ці залежності якісно схожі. Чи не єдиним винятком є поява або зникнення піків перед ділянкою різких змін на залежностях $l(\text{noise})$, $l_w(\text{noise})$ і $M(\text{noise})$. Зауважимо, що відповідні немонотонності функцій $l(\text{noise})$, $l_w(\text{noise})$ і $M(\text{noise})$ утруднюють однозначне розв'язання нашої оберненої задачі — визначення рівня шуму за величиною мережевого параметра.

Корисними є порівняння вивчених мережевих властивостей зображень із відповідними властивостями для ТПМ і рандомізованих текстів, які можна вважати

Таблиця 2. Відмінності кількох мережевих параметрів інформативного ($noise = 0$) та істотно зашумленого ($noise = 60$) зображень для різних радіусів сусідства r .

Table 2. Differences of several network parameters for the informative ($noise = 0$) and essentially noisy ($noise = 60$) images, as obtained at different neighborhood radii r .

Радіус сусідства r	Відмінність мережевих параметрів при рівнях шумів $noise = 60$ і $noise = 0$					
	Абсолютні зміни C	Відносні зміни C , %	Абсолютні зміни l	Відносні зміни l , %	Абсолютні зміни k	Відносні зміни k , %
1	0,120	15,0	-0,258	-19,3	75,3	47,8
2	0,074	8,01	-0,144	-12,6	49,6	24,4
3	0,046	4,85	-0,087	-8,07	36,0	16,5
4	0,034	3,57	-0,060	-5,70	29,5	13,1
5	0,027	2,82	-0,045	-4,33	25,9	11,3

деякими аналогами інформативних і шумових зображень, відповідно. Для простоти обмежимося лише найпростішою моделлю «мішка зі словами», у якій рандомізовані тексти одержують із ТПМ багатократною перестановкою слів [21]. Порівнюючи наші дані з даними, наведеними у праці [21] для багатьох текстів при $r = 2, 3$, бачимо, що закономірності для лінгвістичних мереж і мереж, побудованих на зображеннях, бувають і схожими, і помітно, а іноді навіть радикально відмінними:

1) відмінності окремих мережевих параметрів для ТПМ і рандомізованих текстів незначні або й відсутні: кількість вузлів V однакова, коефіцієнти кластерності C відмінні на 6%, а довжини l найкоротшого шляху – лише на 3%; це помітно відрізняється від ситуації для мереж зображень (порівн. із даними Таблиці 2);

2) при переході від ТПМ до рандомізованих текстів довжина l_w найкоротшого шляху у зваженій мережі зростає на 26%, а для зображень – помітно (на $\sim 25\%$) зменшується (див. Таблицю 1);

3) за тих же умов параметр k лінгвістичних мереж зростає на 18%, а параметр M спадає на 17%, що загалом схоже до ситуації для мереж зображень (див. Таблиці 1 і 2);

4) порівняно з мережами для природних текстів, шляхи l і зважені шляхи l_w для мереж нашого інформативного зображення помітно коротші (відповідно, $l \approx 1,3$ проти 1,8 і $l_w \approx 0,15$ проти 0,3); надзвичайно високими є коефіцієнти кластерності C мереж зображення (0,85 проти $\sim 0,5$);

5) мережеві властивості сильно зашумлених зображень вирізняються найбільшою специфікою: $C \rightarrow 1$, $l \rightarrow 1$, $l_w \rightarrow 1$ і $k \rightarrow V$, що радикально відмінно від властивостей лінгвістичних мереж рандомізованих текстів;

6) як і лінгвістичні мережі текстів, мережі інформативних і шумових зображень є «тісними світами»; якщо спрощено взяти відношення $\mu = C/l$ як параметр «тісноти світу» (див. [22]), ми бачимо, що мережеві «світи», побудовані на інформативних зображеннях, істотно тісніші за «світи» лінгвомереж, а «світи» зображень білого шуму екстремально «тісні» (параметр μ набуває мінімального теоретичного значення $\mu \sim 1$ – порівняно із даними $\mu \sim 0,3$ для мереж рандомізованих текстів). Це вкотре заперечує зв'язки явища «тісного світу» з семантикою тексту або інформативністю зображення (див. висновки [21]).

Нагадаємо означення основних модельних мереж. Це регулярна ґратка (великі l і великі C), ВГ (малі l і малі C) і «тісний світ» (малі l і великі C) [8, 23], причому приклади ще одного гіпотетичного модельного типу мереж (великі l і малі C) відсутні. Отже,

мережа Гаусового шуму – це екстремально «тісний світ», який виявляє дуже малі l і дуже великі C . По-суті це мережа, де кожен вузол зв'язаний з будь-яким іншим вузлом, а всі сусіди кожного вузла є також сусідами одне одного.

На завершення торкнемося опису шляхів практичного розв'язання оберненої задачі – знаходження невідомого рівня шуму в деякому зображенні. На початку зазначимо, що, взагалі кажучи, ми не можемо бути впевненими в тому, що наше початкове досліджуване зображення було повністю вільним від шумів. Тому існує деяка невизначеність абсциси $noise$ у всіх наших залежностях: можливо, її слід замінити на $(noise_0 + noise)$, де $noise_0$ – це шум початкового зображення. Відповідно, використовувати абсолютне значення деякого мережевого параметра для встановлення рівня шуму дещо ризиковано. До того ж, мережеві характеристики $V(noise)$, $C(noise)$, $l(noise)$, $l_w(noise)$, $k(noise)$ і $M(noise)$ для різних початкових зображень можуть бути зміщеними по обох осях абсцис і ординат.

Проте навіть якщо згадані функції матимуть різні початкові та кінцеві ординати зі зростанням шуму для різних зображень, для розв'язання оберненої задачі можна спробувати використати основну рису всіх мережевих характеристик, які ми позначимо як $p(noise)$: функція $p(noise)$ майже незмінна на ділянці низьких шумів, далі різко зростає (або спадає) і нарешті знову набуває майже постійних значень. Тому до досліджуваного зображення з невідомим шумом $noise_0$ слід дозовано додавати шум аж до дуже великих значень $noise$ і спостерігати за поведінкою функції $p(noise_0 + noise)$. Дещо спрощуючи картину, шум $noise_0$ початкового зображення можна оцінити, спостерігаючи, де саме розміщена перша точка $p(noise_0)$ на графіку, і порівнюючи цей графік зі стандартними залежностями на Рис. 3–5. Наприклад, якщо після першої точки $noise_0$ значення $p(noise_0 + noise)$ ще довго будуть порівняно малими і аж згодом почнуть різко змінюватися, то шум $noise_0$ початкового зображення гарантовано був низьким. Дещо важкою задача стане, якщо початкова точка $p(noise_0)$ буде розташована на ділянці різких змін: тоді доведеться визначати рівень шуму $noise_0$, аналізуючи найперше абсолютну величину $p(noise_0)$ і порівнюючи ширину $\Delta(noise)$ одержаної ділянки різких змін із відповідною шириною ділянки різких змін на графіках Рис. 3–5. Звісно, потрібні практичні уточнення такого підходу, суть яких буде найліпше збагнути, виконуючи систематичні дослідження та аналіз великого набору тестових зображень.

5. ВИСНОВКИ

У цій роботі запропоновано новий підхід до аналізу відносного вмісту Гаусового шуму в цифрових зображеннях на основі технічних засобів, запозичених зі статистичної лінгвістики, – визначення мережевих параметрів лінгвістичних мереж. Описано кілька методів побудови незважених і зважених мереж для зображень, а також звернуто увагу на особливості дозованого зашумлення зображень з метою вивчення залежностей мережевих параметрів від відносного рівня $noise$ білого шуму. Виявилось, що для більшості мережевих параметрів p одержуємо монотонно зростаючі або спадні функції $p(noise)$, які типово мають три основні діапазони: дві наближено горизонтальні ділянки при низьких і високих шумах і проміжну ділянку різких змін. Ці особливості, а також помітні відмінності величин p на ділянках малих і великих $noise$ дають змогу впевнено відрізнити інформативні зображення від зображень, які містять значну шумову компоненту – або навіть спробувати більш чи менш строго розв'язати обернену задачу встановлення абсолютної величини початкового рівня шуму в досліджуваному зображенні.

Серед перспектив розвитку запропонованого нами підходу вбачаємо вивчення мереж для зображень, побудованих на основі n -грам значень пікселів, аналіз наслідків альтернативних означень сусідства на зображенні, порівняння мережевих властивостей для великого набору тестових зображень, їхнє порівняння з

властивостями мереж зображень на зразок ВГ, а також дослідження явищ безмасштабності для мереж зображень.

Нарешті, доречним є порівняння мережевого підходу з іншими метриками оцінювання рівня шумів, відомими в галузі опрацювання зображень, – наприклад, ВСШ, піковим ВСШ і дисперсією шуму. Окрім мережевих підходів, ми також плануємо аналізувати застосування до зображень інших методів опрацювання природної мови на зразок вимірювання параметрів повторюваності елементів зображень, а також докладніше вивчати кластеризацію та автокореляції в «текстах», сформованих на основі зображень.

ДОТРИМАННЯ ЕТИЧНИХ СТАНДАРТІВ

Автори заявляють, що дослідження проводилися за відсутності будь-яких конфліктів інтересів.

ВНЕСОК АВТОРІВ

Концептуалізація, [О.Куш.]; методологія, [О.Куш., О.Кр.]; програмне забезпечення, [О.Кр.]; валідація, [О.Кр., В.Ф.]; формальний аналіз, [О.Куш., О.Кр., В.Ф.]; дослідження, [О.Кр., О.Куш.]; матеріали та ресурси, [О.Кр., О.Куш.]; написання – підготовка оригінального тексту, [О.Куш., О.Кр.]; написання – рецензування та редагування, [О.Куш.]; візуалізація, [О.Кр., В.Ф.]; загальний нагляд, [О.Куш.]; адміністрування [О.Куш.].

Усі автори ознайомилися з остаточною версією рукопису та погодилися з його публікацією.

ДЖЕРЕЛА

- [1] Gonzalez R., Woods R. Digital image processing / Pearson, 2018. – 1020 p.
- [2] Clark A., Fox C., Lappin S. The handbook of computational linguistics and natural language processing / Wiley-Blackwell, 2010, – 802 p.
- [3] Meriem H., Farida M. H. Detection of a region of interest in the images based on Zipf laws // In: Proc. 7th Int. Conf. on Signal Image Technol. & Internet-Based Systems. – 2011, – P. 416–421,
- [4] Manzanera A. Σ - Δ background subtraction and the Zipf law / In: Progress in Pattern Recognition, Image Analysis and Applications (CIARP'07). – Viña del Mar-Valparaíso, Chile, 2007. – 10 p. URL: <https://hal.science/hal-01222660/document>
- [5] Довгань І. Я., Кушнір О. С., Фурґала Ю. М., Паночко Г. І. Аналогії між зображеннями та текстами: явище «спалахів» у текстах і цифрових зображеннях // Електроніка та інформаційні технології. – 2022. – Вип. 17. – С. 3–15.
- [6] Zhang Z. Image noise: detection, measurement, and removal techniques / Knoxville: Department of Electrical Engineering and Computer Science, 2015. – 10 pp. URL: <https://zzutk.github.io/docs/reports/2015.07%20-%20Image%20Noise%20Detection,%20Measurement,%20and%20Removal%20Techniques.pdf>
- [7] Olsen S. I. Estimation of noise in images: An evaluation // Graphical Models and Image Processing. – 1993. – Vol. 55, No. 4. – P. 319–323.
- [8] Newman M. E. J. The structure and function of complex networks // SIAM Rev. – 2003. – Vol. 45. – P. 167–256.
- [9] Ferrer i Cancho R., Solé R. V. The small world of human language // Proc. Roy. Soc. Lond. B. – 2001, – Vol. 268. – P. 2261–2265.
- [10] Dorogovtsev S. N., Mendes J. F. F. Language as an evolving word web // Proc. Roy. Soc. Lond. B. – 2001, – Vol. 268. – P. 2603–2606.

- [11] Matsuo Y., Ohsawa Y., Ishizuka M. KeyWorld: Extracting keywords from a document as a small world / In: Jantke K. P., Shinohara A. (Eds.). *Discovery Science*, 2001, Lecture Notes in Computer Science, Vol. 2226. – Berlin, Heidelberg: Springer.
- [12] Ferrer i Cancho R., Solé R. V., Köhler R. Patterns in syntactic dependency networks // *Phys. Rev. E*. – 2004. – Vol. 69. – P. 051915 (8 pp.).
- [13] Antiqueira L., Nunes M. G. V., Oliveira Jr. O. N., Costa L. da F. Strong correlations between text quality and complex networks features // *Physica A*. – 2007. – Vol. 373. – P. 811–820.
- [14] Cong J., Liu H. Approaching human language with complex networks // *Phys. Life Rev.* – 2014. – Vol. 11. – P. 598–618.
- [15] Espitia D., Larralde H. Universal and non-universal text statistics: Clustering coefficient for language identification // *Physica A*. – 2020. – Vol. 553. – P. 123905 (25 pp.).
- [16] Masucci A. P., Rodgers G. J. Network properties of written human language // *Phys. Rev. E*. – 2006. – Vol. 74. – 026102 (8 pp.).
- [17] Caldeira S. M. G., Petit Lobão T. C., Andrade R. F. S., Neme A., Miranda J. G. V. The network of concepts in written texts // *Eur. Phys. J. B*. – 2006. – Vol. 49. – P. 523–529.
- [18] Masucci A. P., Rodgers G. J. Differences between normal and shuffled texts: Structural properties of weighted networks // *Adv. Complex Syst.* – 2009. – Vol. 12. – P. 113–129.
- [19] Amancio D. R., Altmann E. G., Rybski D., Oliveira Jr. O. N., Costa L. da F. Probing the statistical properties of unknown texts: Application to the Voynich manuscript // *PLOS ONE*. – 2013. – Vol. 8. – e67310 (10 pp.).
- [20] Carretero-Campos C., Bernaola-Galván P., Ivanov P. Ch., Carpena P. Improving statistical keyword detection in short texts: entropic and clustering approaches // *Physica A*. – 2013. – Vol. 392. – P. 1481–1492.
- [21] Кушнір О. С., Дребот А. С., Д. А. Остріков, Кравчук О. М. Властивості лексичних мереж, побудованих на природних і рандомних текстах // *Електроніка та інформаційні технології*. – 2024 – Вип. 28. – С. 22–37.
- [22] Neal Z. P. How small is it? Comparing indices of small worldliness // *Network Science*. – 2017. – Vol. 5, No. 1 – P. 30–44.
- [23] Börner K., Sanyal S., Vespignani A. Network science / In: B. Cronin (Ed.). *Annual Review of Information Science & Technology*, Vol. 41, Medford, NJ: Information Today / American Society for Information Science and Technology. – Chapter 12. – pp. 537–607.

PROPERTIES OF NETWORKS BUILT ON DIGITAL IMAGES

Oleh Kushnir^{*}, Oleksiy Kravchuk^{*}, Volodymyr Franiv^{*}

Department of Optoelectronics and Information Technologies,
Ivan Franko National University of Lviv
107 Tarnavsky Street, UA-79017 Lviv, Ukraine

ABSTRACT

Background. It is known that some important properties of natural-language texts can be revealed when mapping a text on a network and studying the properties of this network. Here the words of a text are regarded as nodes and their co-occurrences as links.

Materials and Methods. Drawing an analogy between texts and digital images, in this work we study experimentally and analyze phenomenologically the properties of a network

built on an image. In this network, the nodes are pixel values and the links are assigned among those pixel values that spatially adjoin each other in an image. The neighborhood of pixels is defined by a radius r ($r = 1, 2, \dots$). We build our networks for a test informative image and the corresponding images obtained from the initial one by adding calibrated portions of white Gaussian noise. Both weighted and non-weighted networks are analyzed and a number of practical methods for building an image network and adding a noise to an image are compared with each other.

Results and Discussion. The main network parameters such as the average clustering coefficient and the average shortest-path length are measured as functions of the relative noise parameter. The main qualitative and quantitative features of the dependences are analyzed. The network characteristics are critically compared with those known for the lexical networks build upon the natural texts and the random texts obtained with randomizing words in the initial natural text. The main similarities and distinctions of the lexical and image networks are scrutinized.

Conclusion. It is shown that the average clustering coefficient and path length for the networks built upon the initial informative image and the corresponding noisy images are essentially small worlds. In spite of this similarity, the above networks reveal a number of distinct features so that the dependences of the image-network parameters on the noise level can be used for distinguishing between these types of images. Finally, we analyze a number of consequences of our empirical results and some data known from the literature. Since both informative and noisy images reveal a small-worldliness, it would hardly be appropriate to associate this effect with the information load of the image.

Keywords: complex networks, analysis and classification of images, image recognition, information and noise detection, semantics, random models.

UDC 004.93

ENHANCEMENT OF SENSOR PANEL TACTILE TOUCH INTERFACE

Oleksandr Karpin^{1,2}, Zinovii Liubun¹, Vasyl Mandziy^{1,2},
Oleh Tereshchuk¹, Nestor Hotsiy^{2,3}

¹ Department of RadioPhysics and Computer Technologies,
Department of Sensor and Semiconductor Electronics,
Ivan Franko National University of Lviv
107 Tarnavsky Str., 79017 Lviv, Ukraine

² Infineon Technologies, Lviv, Ukraine,
20 Luhanska Str., 79048 Lviv, Ukraine

³ Department of Information Technologies and Electronic Communications Systems,
Lviv State University of Life Safety,
35 Kleparivska Str., 79007 Lviv, Ukraine

Karpin, O. et al. (2025). Enhancement of sensor panel tactile touch interface. Electronics and Information Technologies, 29, 29-36. <https://doi.org/10.30970/eli.29.3>

ABSTRACT

Introduction. The article explores the possibility of utilizing a touch-sensitive button for the identification of two types of signals – a button press signal and a finger swipe signal across the button. This concept has the potential to revolutionize the way we interact with electronic devices, making them more intuitive and user-friendly. The ability to detect different types of gestures using a single button can also lead to a more compact and cost-effective design.

Materials and Methods. To address the classification task, an algorithm has been proposed for the identification of signal characteristics upon which recognition will be conducted. The algorithm is based on a neural network architecture, which is trained on a dataset of signals collected from a touch-sensitive button. The dataset includes a variety of signals, each corresponding to a specific gesture, such as a button press or a swipe in different directions. The neural network is designed to learn the patterns and characteristics of the signals, allowing it to accurately classify new, unseen signals. The algorithm is optimized to minimize processing time and computational resources, making it suitable for real-time applications.

Results. Using neural networks to solve the recognition task allows for the easy determination of optimal classification algorithm parameters for a specific type of touch button. The results show that the proposed algorithm achieves high accuracy in identifying two types of gestures, with a minimal error rate.

Conclusions. The proposed classification algorithm exhibits satisfactory accuracy in identifying the two signals within a minimal timeframe and requires minimal computational resources. Therefore, it can be employed cost-effectively to enhance the functionality of touch panels. The algorithm's ability to detect gestures using a single button makes it an attractive solution for applications where space is limited, such as in wearable devices or mobile phones. Additionally, the algorithm's low computational requirements make it suitable for use in low-power devices, such as those powered by batteries or energy harvesting systems. Future work can focus on improving the algorithm's accuracy and robustness, as well as exploring its application in different domains.

Keywords: gesture detection, neural network, capacitive sensor, signal profile.



© 2025 Oleksandr Karpin et al. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Using gestures is one of the most convenient ways of interaction with gadgets. Therefore, gesture real time detection has many applications [1-6], including gaming, virtual reality, and smart home devices. The ability to detect gestures in real-time enables users to interact with devices in a more natural and intuitive way, enhancing the overall user experience.

Development of structures and algorithms of such gadgets combined with the advantages of microcontrollers Infineon Technologies [7] and neural networks shall allow realizing simple, cheap, effective, and reliable gadgets. The integration of Infineon's microcontrollers with neural networks can provide a powerful platform for gesture detection, enabling the development of a wide range of innovative applications.

The use of neural networks can improve the accuracy and robustness of gesture detection, while Infineon's microcontrollers can provide the necessary processing power and low power consumption. This combination can enable the creation of gesture-based devices that are not only effective but also energy-efficient and cost-effective.

MATERIALS AND METHODS

Figure 1 shows a panel prototype with three buttons. Only signal from one of them (middle button) was used for the algorithm development. A button press is identified as a certain command.

An algorithm shall be developed to differentiate the two types of signal:

1. The sensor signal resulted from a button press (Touch).
2. The sensor signal resulted from a finger swiping on the button surface (Swipe).

To correctly classify signals, signal attributes are required. Figure 2a and 2b shows pairs of real-signals examples that differ visually while Fig. 2c and 2d – similar examples that can hardly be differed visually.

The requirements to the identification system productivity: the classification shall be executed immediately after the signal start and ensure the algorithm high efficiency. Therefore, for example, the signal spectrum analysis cannot be used due to its inherent features:

- the need to process the entire signal (from start to end);
- significant amount of processing time;
- significant amount of calculation resources for the transition to the frequency domain.

So, simple attributes are required. Neither the signal level nor its duration is applicable. They may change in a wide range for both signal types. Only the signal shape is applicable.

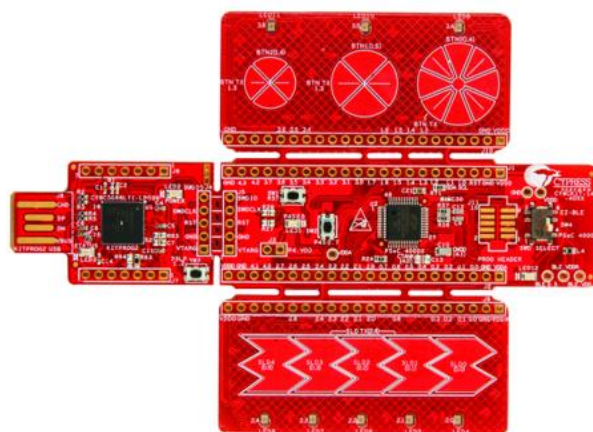


Fig. 1. PSoc 4000S Prototyping Board

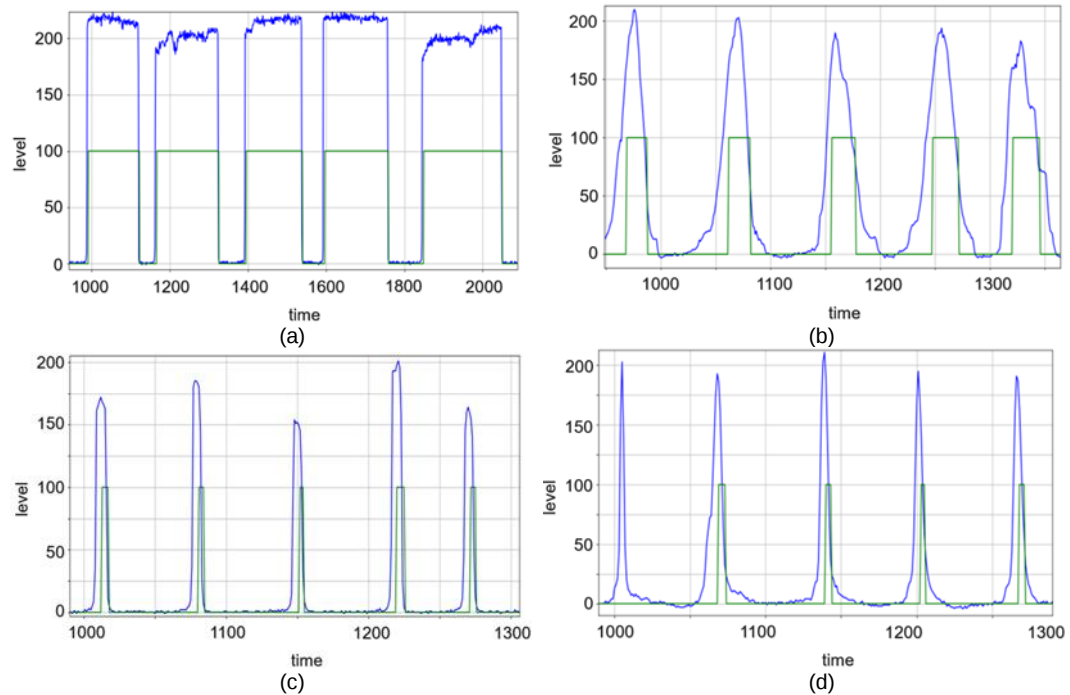


Fig. 2. Real signals: long touch (a); slow right to left swipe (b); short touch (c) and regular right to left swipe (d). Signals are depicted by blue lines, whereas device-identified results – by green lines.

Touch-type signal is close to the Π -shaped in most cases. Swipe-type signal is more similar to the Gauss function.

Unfortunately, short-duration signals in fact cannot be differentiated even visually (Fig. 2c and 2d). This leads to the consideration that any identification algorithm will not yield a 100% correct result. In such cases, the classification issue can be resolved with the neural network solution.

Similar to any classification algorithms, the neural network approach requires the selection of the classification attributes. As mentioned above, most signals can be differentiated based on their shape and the signal derivative depends significantly on the shape (see Figures 3, 4). Therefore, the analysis of the signal shape transits to the analysis of values of the signal derivative, where the derivative is the time function.

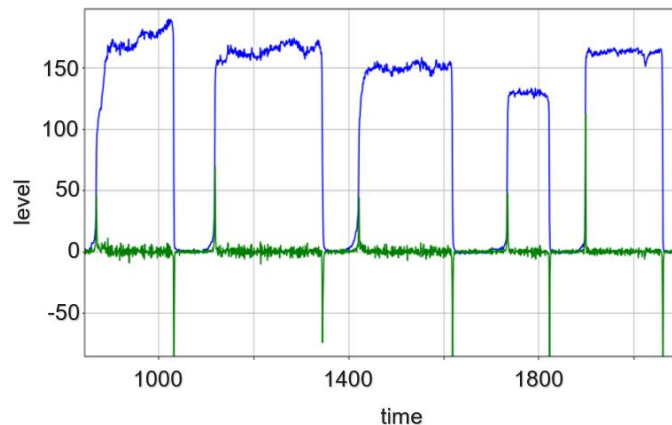


Fig. 3. Long touch signals. Blue line – real signal, green line – signal derivative.

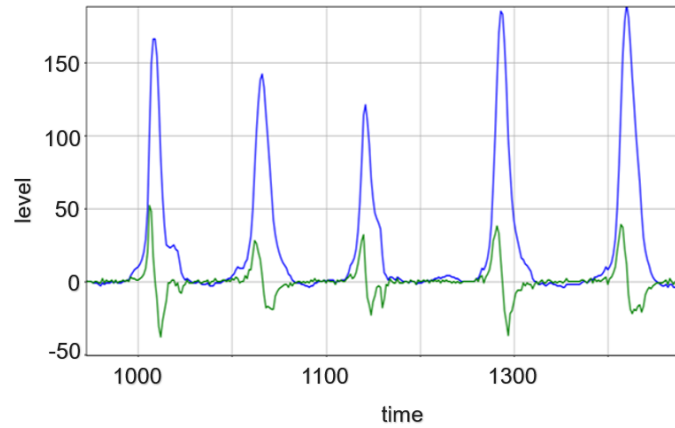


Fig. 4. Regular left to right swipe signals. Blue line – real signal, green line – signal derivative.

The Π -shaped signal may appear suddenly between two peaks determined by the edges of the time interval signal for which the derivative is close to zero. For the Swipe type of signal, such an interval is either absent or minimal. This difference can be used as one of the attributes for the classification. Another attribute can be the speed of the increasing of the leading-edge signal. In many cases, this speed is significantly bigger for the Touch type signal than for the Swipe type signal. Therefore, the maximal value of the derivative on the leading-edge signal can be taken as the second attribute.

RESULTS AND DISCUSSION

The two attributes are selected for the classification:

1. The maximal value of the derivative on the leading-edge signal.
2. The number of zeros on the signal top.

Figure 5 shows the flowchart for signal attributes determination.

To make the algorithm work, the two constants are required: the signal minimal level, the maximal value of the signal derivative, where the derivative shall be equal to zero.

Figure 6a demonstrates the results obtained by the program for button press (Touch signal). Figure 6b – same results for button surface (Swipe signal). Figure 6 confirms that the attributes values are different. On the other hand, there is a significant overlapped area (see Fig. 7).

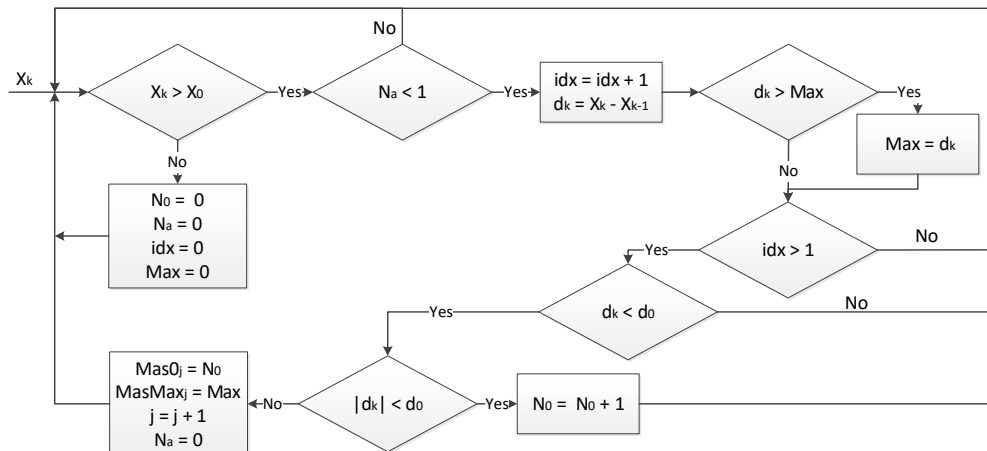


Fig. 5. Algorithm for determination of signal attributes.

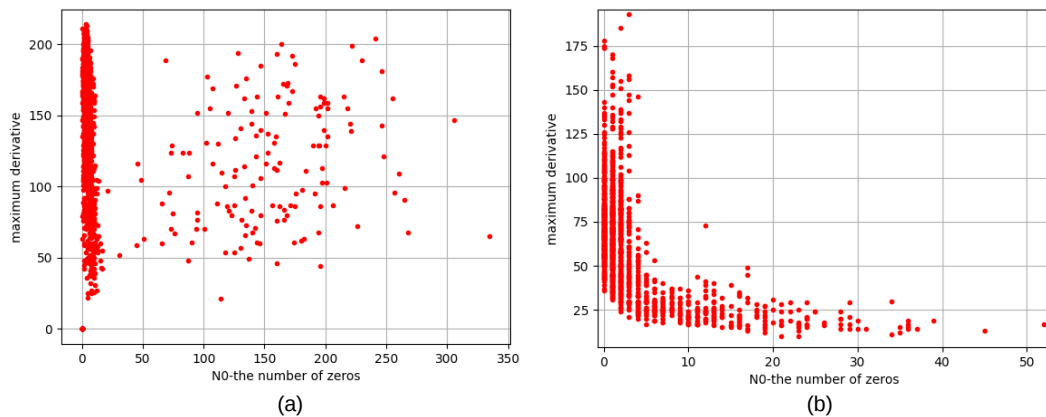


Fig. 6. Signals attributes values for Touch signal (a) and Swipe signal (b).

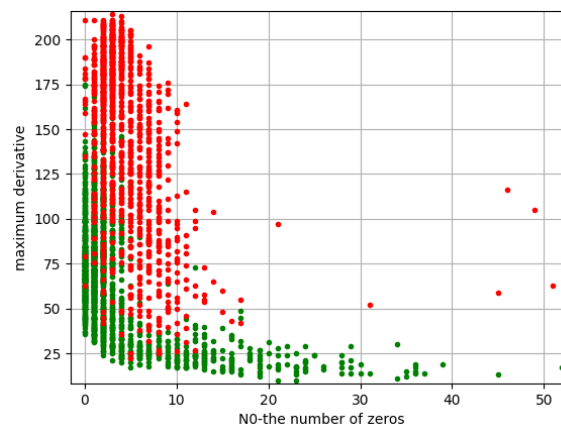


Fig. 7. Normalized attributes values for two types of signals: green dots – Swipe, red dots – Touch.

Due to the difficulty of the signal's boundary determination, as the solution for the classification task, it was selected the forward propagation two-layer neural network with the sigmoidal function of activation, which structure is depicted on the Fig. 8.

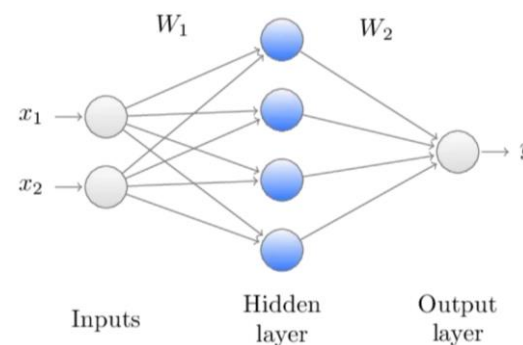


Fig. 8. Neural network structure.

Before the neural network training, the input vectors were normalized. The number of neurons in the input layer shall determine the accuracy of reproducing the boundary between the two sets. To assess the quality of the classification, the error coefficient K was used:

$$K = (ER_{swipe} + ER_{touch})/N_{np}, \quad (1)$$

where:

ER_{swipe} is the number of incorrectly determined signals of the Swipe type;

ER_{touch} – the number of incorrectly determined signals of the Touch type;

N_{np} – the summative number of signals.

Table 1 contains the results of the classification with different quantities of neurons in the input layer.

The further increase of the number of neurons in the input layer shall not yield better results because the boundary between the two types of signals is blurry. The only way to improve the classification accuracy is to increase the number of the classification attributes. The selected here attributes worked the best. Whereas, additional attributes, for example, the signal amplitude value did not show significant improvement.

Table 1. Results of the classification with different quantities of neurons in the input layer

#	Number of neurons in the input layer	Error coefficient (%)
1	2	8.2 – 9.0
2	3	7.1
3	4	6.5
4	6	6.4
5	8	6.3

Figure 9 shows the offset limit of the signals obtained by the trained neural network with six neurons in the input layer.

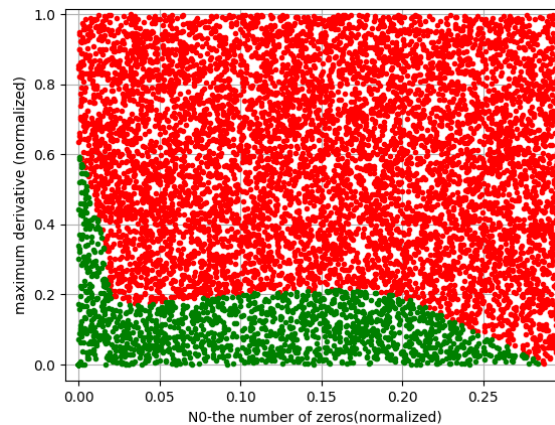


Fig. 9. Boundary between the signals (green dots – Swipe signal, red dots – Touch signal).

The further increase of the number of neurons in the input layer shall not yield better results because the boundary between the two types of signals is blurry. The only way to improve the classification accuracy is to increase the number of the classification attributes.

The selected here attributes worked the best. Whereas, additional attributes, for example, the signal amplitude value did not show significant improvement.

CONCLUSION

Based on obtained results, we have concluded that the proposed classification algorithm yields quite good accuracy while executing the identification of two signals in the minimal time period and requires minimal calculation resources, so, can be used with minimal cost for the quality enhancement of the sensor panels.

COMPLIANCE WITH ETHICAL STANDARDS

The author declare that the research was conducted in the absence of any potential conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, [O.K.]; methodology, [Z.L.]; validation, [O.K.]; formal analysis, [V.M.]; investigation, [O.T.]; resources, [Z.L.]; data curation, [Z.L.]; writing – original draft preparation, [Z.L.]; writing – review and editing, [V.M.]; visualization, [N.H.] supervision, [V.M.]; project administration, [O.K.]; funding acquisition, [O.K.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Z. Liubun, V. Mandziy, O. Karpin, V. Rabyk, "Correlation-based Contactless Hand-Gesture Detection for Capacitive Sensing". 2021 IEEE 16th International Scientific and Technical Conference "Computer Science and Information Technologies (CSIT)", 2022, DOI: 10.1109/TCSET55632.2022.9767083.
- [2] Klein H., Karpin O., Kravets I., Kolych I., MacSweeney D., Ogirko R., O'Keefe D., Walsh P. "Advanced Capacitive Sensing for Mobile Devices" // Low-Power Analog Techniques, Sensors for Mobile Devices, and Energy Efficient Amplifiers. Springer, 2019, pp 131–153, https://doi.org/10.1007/978-3-319-97870-3_7
- [3] Aldi Sidik Permana, Esmeralda Contessa Djamal, Fikri Nugraha, Fatan Kasyidi, "Hand Movement Identification Using Single-Stream Spatial Convolutional Neural Networks", Electrical Engineering Computer Sciences and Informatics (EECSI) 2020 7th International Conference, pp. 172-176, 2020. <https://journal.portalgaruda.org/index.php/EECSI>
- [4] Heng-Tze Cheng, An Mei Chen, Ashu Razdan, Elliot Buller, "Contactless Gesture Recognition System Using Proximity Sensors", IEEE International Conference on Consumer Electronics (ICCE), 2020 <https://ieeexplore.ieee.org/document/5722510>
- [5] Alghamdi S., Elrefaei L. Dynamic user verification using touch keystroke based on medians vector proximity. In Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference, pp. 121– 126. IEEE, 2015
- [6] Abisado M., Gerardo B., Fajardo A. Towards keystroke analysis using neural network for multi - factor authentication of learner recognition in on - line examination. In Manila International Conference on Trends in Engineering and Technology (2017), pp. 71–74
- [7] Infineon Technologies and CY8CKIT-040T PSoC™ 4000T CAPSENSE™ Evaluation Kit <https://www.infineon.com/cms/en/product/evaluation-boards/cy8ckit-040t/> , 2024

АНАЛІЗ УДОСКОНАЛЕННЯ ТАКТИЛЬНОГО ІНТЕРФЕЙСУ ДОТИКУ СЕНСОРНОЇ ПАНЕЛІ

**Олександр Карпін^{1,2}, Зіновій Любунь¹, Василь Мандзій^{1,2},
Олег Терещук¹, Нестор Гоцій^{2,3}**

¹ Кафедра радіофізики та комп'ютерних технологій
Кафедра сенсорної та напівпровідникової електроніки,
Львівський національний університет імені Івана Франка
вул. Тарнавського, 107, м. Львів, 79017, Україна

² Інфінеон Технолоджіс, Львів, Україна
вул. Луганська 20, м. Львів, 79048, Україна

³ Кафедра інформаційних технологій та систем електронних комунікацій
Львівський державний університет безпеки життєдіяльності
вул. Клепарівська 35, м. Львів, 79007, Україна

АНОТАЦІЯ

Вступ. В статті розглянуто можливість використання сенсорної кнопки для ідентифікації двох типів сигналів – сигналу від натискання кнопки та сигналу від проведення пальцем по кнопці. Ця концепція має потенціал кардинально змінити спосіб взаємодії з електронними пристроями, зробивши їх більш інтуїтивно зрозумілими та зручними для користувача. Здатність виявляти різні типи жестів за допомогою однієї кнопки також може призвести до більш компактного та економічно ефективного дизайну.

Матеріали та методи. Для вирішення задачі класифікації запропоновано алгоритм визначення ознак сигналів на основі яких буде проводитись розпізнавання. Алгоритм базується на використанні нейронної мережі, яка навчається на наборі даних сигналів, зібраних із сенсорної кнопки. Набір даних містить різноманітні сигнали, кожен з яких відповідає певному жесту, наприклад натискання кнопки або проведення пальцем у різних напрямках. Нейронна мережа розроблена для вивчення шаблонів і характеристик сигналів, що дозволяє їй точно класифікувати нові, невідомі сигнали. Алгоритм оптимізовано для мінімізації часу обробки та обчислювальних ресурсів, що робить його придатним для детектування жестів в режимі реального часу.

Результати. Використання нейромереж для вирішення задачі розпізнавання дозволяє легко визначити оптимальні параметри алгоритму класифікації для конкретного типу сенсорної кнопки. Результати показують, що запропонований алгоритм забезпечує високу точність ідентифікації двох типів жестів з мінімальною кількістю помилок.

Висновки. Запропонований алгоритм класифікації дає досить добру точність при виконанні ідентифікації двох сигналів за мінімальний проміжок часу та вимагає мінімальних обчислювальних ресурсів. Таким чином, він може бути використаний з мінімальними витратами для збільшення функціональності сенсорних панелей. Здатність алгоритму виявляти жести за допомогою однієї кнопки робить його привабливим рішенням для додатків, де простір обмежений, наприклад, у переносних пристроях або мобільних телефонах. Крім того, низькі обчислювальні вимоги алгоритму роблять його придатним для використання в малопотужних пристроях, наприклад тих, що живляться від батареї. Майбутня робота може бути зосереджена на покращенні точності та стійкості алгоритму, а також на дослідженні його застосування в різних областях.

Ключові слова: Ємнісний сенсор, профіль сигналу, індикація, нейронна мережа.

УДК 004.6, 004.8

ВИЗНАЧЕННЯ НАЙКРАЩОГО ОПТИМІЗАТОРА ДЛЯ НЕЙРОМЕРЕЖІ ПРИ РОЗРОБЦІ СИСТЕМ АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ

Андріан Козинець 

Львівський національний університет імені Івана Франка,
вул. Драгоманова 50, 79005 м. Львів, Україна

Козинець, А. (2025). Визначення найкращого оптимізатора для нейромережі при розробці систем автоматичної класифікації зображень. *Електроніка та інформаційні технології*, 29, 37-46. <https://doi.org/10.30970/eli.29.4>

АНОТАЦІЯ

Передумови. Вибір найкращого оптимізатора є важливим кроком у розробці ефективних систем автоматичної класифікації зображень. Зокрема, для нейромереж, побудованих на основі згорткових нейронних мереж (ЗНМ), вибір між популярними методами оптимізації, такими як Adam (Адаптивна Оцінка Моменту) та SGD (Стохастичний Градієнтний Спуск), може суттєво вплинути на їхню ефективність. Порівняння цих оптимізаторів дає змогу визначити найбільш відповідний оптимізатор для розв'язання конкретних завдань машинного навчання.

Методи та інструменти. Для навчання та тестування моделей було використано бібліотеку TensorFlow та інтерфейс для візуалізації процесу машинного навчання TensorBoard. Нейромережі тренувались на двох різних наборах даних з суттєво різними характеристиками. Для кожного набору даних було проведено навчання за допомогою двох оптимізаторів: Adam та SGD. Під час аналізу результатів враховувалися такі параметри, як точність тренування та перевірки, а також втрати під час тренування та перевірки.

Результати та обговорення. Згідно з результатами, оптимізатор SGD показав кращі показники продуктивності порівняно з Adam. Початкова точність для SGD була вищою, а також спостерігалось значне підвищення швидкості зростання точності. Крім того, остаточна точність моделі з оптимізатором SGD була вищою як на тренувальному, так і на перевірному наборах даних, що вказує на більш ефективний процес навчання з використанням цього оптимізатора.

Висновки. Визначено, що використання оптимізатора SGD призводить до кращих результатів у порівнянні з оптимізатором Adam для задач автоматичної класифікації зображень. Це підтверджується вищими показниками точності та більш швидким процесом навчання.

Ключові слова: TensorFlow, згорткова нейронна мережа, Adam, адаптивна оцінка моменту, SGD, стохастичний градієнтний спуск, GPU, TPU.

ВСТУП

У сучасному цифровому світі обробка та класифікація зображень стали необхідними складовими багатьох додатків та сервісів. З розвитком технологій машинного навчання з'являються нові можливості для автоматизації цих процесів, що значно підвищує ефективність та точність роботи з великими обсягами візуальних даних. Одним із перспективних напрямків є розробка систем автоматичної класифікації зображень, які дозволяють користувачам швидко отримувати інформацію про зміст зображень та полегшують їх пошук і категоризацію.



© 2025 Andrian Kozynets. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Класифікація зображень є процесом асоціювання ключових слів або міток з зображеннями для того, щоб описати їхній зміст. Класифікація має велике значення у різних сферах та застосуваннях, оскільки допомагає в управлінні, пошуку та аналізі зображень. Наприклад, класифікація зображення використовується у соціальних мережах, системах керування вмістом (СКВ), пошукових системах. Під час навчання моделей машинного навчання класифіковані зображення використовуються для покращення алгоритмів розпізнавання об'єктів, сцен та інших характеристик зображень. Це дозволяє створювати більш точні та ефективні системи розпізнавання.

Вибір оптимізатора під час навчання ЗНМ для класифікації зображень є надзвичайно важливим етапом, який безпосередньо впливає на час навчання, точність отриманої моделі та розбіжність між фактичними та передбаченими результатами. В статті [1] розглянуто причину через яку SGD погано працює з проблемою блочної неоднорідності, а також показано чому адаптивна оцінка моменту (Adam) показує кращі результати в цій ситуації. У [2] розглянуто гібридне використання цих двох оптимізаторів, що призводить до отримання кращих результатів навчання.

При виборі оптимізатора для навчання ЗНМ важливо враховувати специфіку задачі та характеристики даних, а також особливості архітектури самої мережі. Однією з сучасних тенденцій є використання методів мета-оптимізації, які автоматично налаштовують параметри навчання. Це значно спрощує процес пошуку найкращих гіперпараметрів і забезпечує стабільнішу та швидшу збіжність. Наприклад, байесова оптимізація може ефективно знаходити оптимальні параметри, комбінувати різні підходи до навчання.

Методи, які комбінують SGD з адаптивними методами, такими як Adam, стають все популярнішими. Такі гібридні оптимізатори можуть знаходити більш ефективні траєкторії мінімізації функції втрат та забезпечувати баланс між швидкістю збіжності та стабільністю. Використання таких оптимізаторів знижує ризик застрягання в локальних мінімумах і покращує узагальнення на нових даних.

Окрім вибору оптимізатора, необхідно враховувати особливості апаратного забезпечення. Наприклад, сучасні графічні процесори (ГП) та тензорні процесори (ТП) значно прискорюють процес навчання завдяки своїй архітектурі, яка підходить для паралельної обробки даних. Оптимізатори, такі як Adam, добре працюють з великими наборами даних і можуть ефективно використовувати обчислювальні ресурси ГП/ТП.

Використання гібридних оптимізаторів і методів мета-оптимізації вимагає системного підходу до налаштування та тестування моделі. Це дозволяє максимально ефективно використовувати доступні обчислювальні ресурси, досягати високої точності та узагальнюваності моделей на нових даних.

МЕТОДИ ТА ІНСТРУМЕНТИ

Для розробки системи автоматичної класифікації зображень використано бібліотеку TensorFlow та два оптимізатори: Adam і SGD.

Вибір бібліотеки TensorFlow зумовлений її широким застосуванням у дослідженнях та промислових проектах, а також наявністю великої кількості інструментів і ресурсів для розробки та оптимізації моделей машинного навчання. TensorFlow надає гнучкість у налаштуванні моделей та підтримує роботу на різних платформах, що робить її зручною для використання в різних середовищах.

Оптимізатор Adam обрано через його здатність адаптивно змінювати швидкість навчання для кожного параметра окремо, що прискорює збіжність моделі та покращує її стабільність. Adam добре підходить для задач, де дані мають високу варіативність.

Оптимізатор SGD, з іншого боку, є класичним методом оптимізації, який часто використовується через свою простоту та ефективність у задачах з великими наборами даних. Він менш вибагливий до ресурсів і може забезпечити хороші результати при правильному налаштуванні швидкості навчання. Проте, порівняно з Adam, SGD може демонструвати повільнішу збіжність і більшу чутливість до вибору початкових гіперпараметрів.

У порівнянні з іншими популярними оптимізаторами, такими як RMSprop або Adagrad, вибір Adam і SGD обґрунтований їхньою здатністю забезпечити хороший баланс між швидкістю навчання та стабільністю моделі. RMSprop, наприклад, теж є адаптивним оптимізатором, але Adam об'єднує в собі переваги RMSprop та моменту, що дає йому перевагу в багатьох сценаріях.

Оптимізатори в 3НМ відіграють ключову роль у процесі навчання. Вони допомагають мережі адаптуватися до набору даних та підвищувати ефективність у вирішенні поставленої задачі. Два найпоширеніші оптимізатори: Adam та SGD – мають свої унікальні особливості та застосування.

Оптимізатор Adam [3] використовує метод ефективної стохастичної оптимізації, який вимагає лише градієнтів першого порядку з невеликими вимогами до пам'яті. Метод розраховує індивідуальні адаптивні темпи навчання для різних параметрів з оцінок першого та другого моментів градієнтів.

SGD — це оптимізаційний алгоритм, який лежить в основі багатьох моделей глибокого навчання (ГН). SGD працює шляхом поступового коригування параметрів моделі з використанням градієнта функції втрат, який обчислюється для підмножини (міні-пакета) даних навчання, а потім змінюючи міні-партію протягом певного часу, щоб вивчити весь набір даних [4].

Для дослідження ефективності оптимізаторів послідовно використано два набори даних. Перший набір даних містив 25 тисяч зображень розділених на 6 категорій. Другий набір даних містив 550 тисяч зображень, розділених на тисячу класів. Кожне зображення у наборах було попередньо підготовлено для використання.

Дослідження проводилося на п'ятишаровій нейронній мережі, яка включала три згорткових шари з ядром розміром 3x3 та попередньою активацією за допомогою функції активації ReLU (Rectified Linear Unit, Випрямлений Лінійний Вузол). Кожний згортковий шар налаштовувався для виявлення різних ознак та характеристик у вхідних зображеннях. Після згорткових шарів використовувався шар згладжування для згладжування вихідних даних перед їх подачею на повнозв'язний шар. Останній, п'ятий, шар мережі був повнозв'язний та використовувався для класифікації зображень на тисячу різних класів, що представляли собою категорії зображень. Темп навчання для обох оптимізаторів встановлено на однаковий рівень, який становив 10^{-4} . Однаковий темп забезпечує порівнянність результатів та дозволяє уникнути впливу різниці у темпі навчання на ефективність оптимізації.

Процес тренування моделі виконувався за допомогою бібліотеки TensorFlow на апаратному забезпеченні, що включає графічну карту RTX 3060 Ti з GDDR6X 8 Gb пам'яті, процесор Ryzen 5 5600 та 16 Гб оперативної пам'яті. Тренування моделі включало кілька етапів: початковий запуск моделі з початковими параметрами, проходження через набір тренувальних даних, обчислення втрат (loss), оновлення параметрів моделі для зменшення цих втрат.

Для кожної епохи модель проходила через весь набір тренувальних даних, і для другого набору даних процес займав приблизно 40 хвилин на кожну епоху. Протягом кожної епохи модель обчислювала втрати на тренувальних даних і коригувала свої параметри для мінімізації цих втрат. Після кожної епохи модель також оцінювалася на перевірному наборі даних, щоб перевірити, як добре вона узагальнює нові дані, які не були використані під час тренування. Загалом було виконано 9 епох тренування, що зайняло приблизно 7 годин 20 хвилин. Під час цього процесу

спостерігалось поступове зменшення втрат як на тренувальному, так і на перевірному наборах даних, що свідчить про ефективне навчання моделі. Результати були візуалізовані у вигляді графіків та таблиць.

Для збору та аналізу результатів навчання використовувався інструмент TensorBoard, який надає можливість візуалізації метрик навчання, таких як втрата та точність, на різних етапах процесу навчання. Це дозволяє систематично оцінювати та порівнювати продуктивність стохастичного градієнтного спуску та адаптивної оцінки моменту під час навчання моделі з метою визначення найефективнішого оптимізатора для даної задачі класифікації зображень.

На рис.1 представлено приклад декількох зображень з набору даних (ambulance, cup, strawberry) та частину назв різноманітних категорій даних.

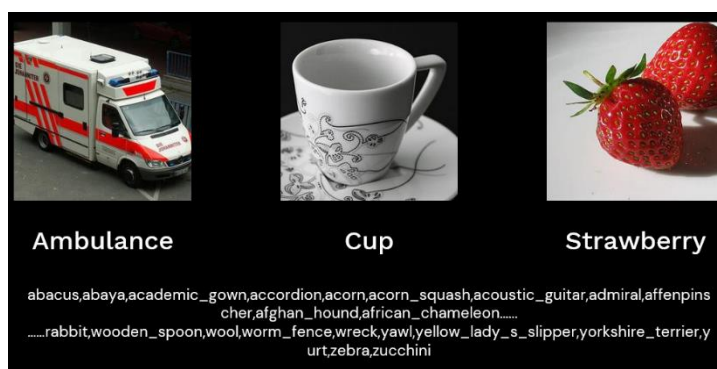


Рис.1. Зображення та назви категорій у наборі даних.

Fig. 1. Images and names of categories in the dataset.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

На рис.2 зображено графіки результатів тренування моделі з використанням оптимізаторів SGD та Adam. Графіки зверху показують результати для оптимізатора SGD. Ліворуч (рис.2а) відображена точність (accuracy) моделі, а праворуч (рис.2б) — втрати (loss) моделі по епохах для тренувальних та перевірних даних. Пунктирні представляють фактичні значення, тоді як суцільні лінії показують згладжені значення. Для оптимізатора SGD точність тренувальних даних поступово зростає і досягає згладженого значення 0,8297 та фактичного значення 0,8343 після 16 епох. Точність перевірних даних також зростає до згладженого значення 0,8141 і фактичного значення 0,8117. Втрати на тренувальних даних зменшуються до згладженого значення 0,4694 і фактичного значення 0,4545, а втрати на перевірних даних зменшуються до згладженого значення 0,541 і фактичного значення 0,5498. Загальний час тренування для кожної епохи становив 11,75 хвилин. Для оптимізатора Adam (графіки внизу) точність тренувальних даних зростає до згладженого значення 0,8218 та фактичного значення 0,8252 після 19 епох.

Точність моделі на перевірних даних зростає до згладженого значення 0,7929 і фактичного значення 0,805. Втрати на тренувальних даних зменшуються до згладженого значення 0,4901 і фактичного значення 0,48, а втрати на перевірних даних зменшуються до згладженого значення 0,5812 і фактичного значення 0,5479. Загальний час тренування для кожної епохи становив 13,91 хвилин. Порівнявши результати двох оптимізаторів, можна побачити, що оптимізатор SGD досягає трохи кращої точності на перевірних даних, тоді як оптимізатор Adam показує трохи менші втрати на перевірних даних. Виходячи з цих результатів, можна зробити висновок, що оптимізатор SGD є дещо кращим варіантом для цієї задачі, оскільки забезпечує вищу точність на перевірних даних.

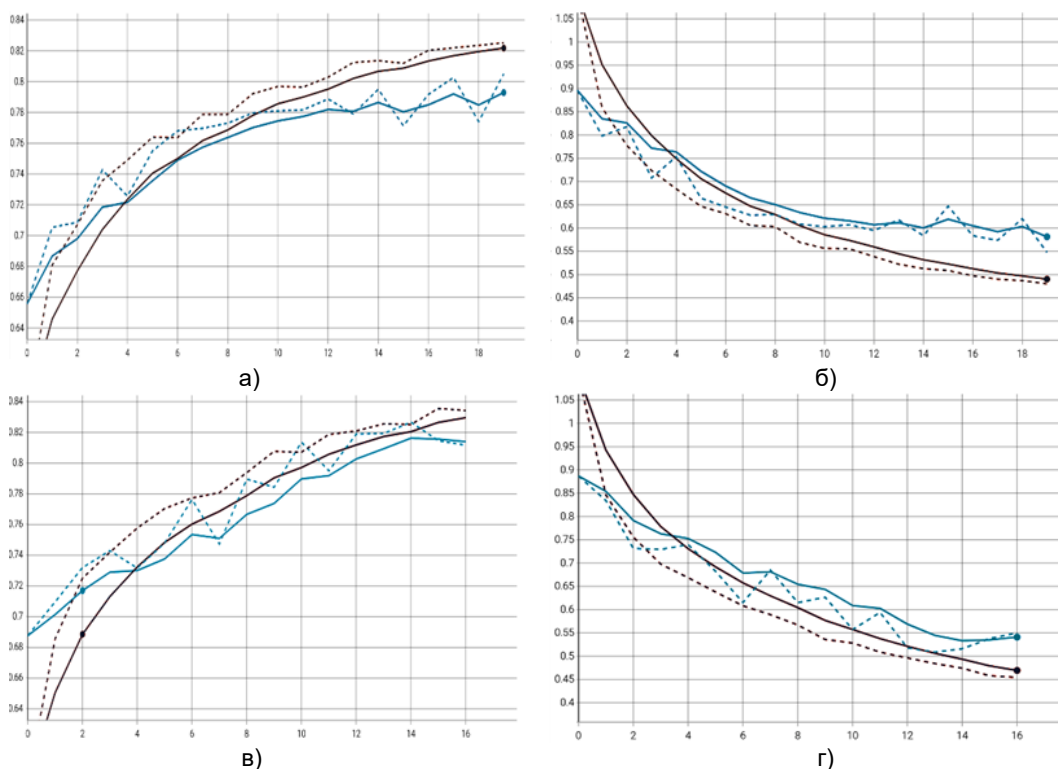


Рис.2. Графіки залежності точності та втрат від кількості епох навчання (перший набір даних): точність оптимізатора SGD (а), втрати оптимізатора SGD (б), точність оптимізатора Adam (в), втрати оптимізатора Adam (г). Коричневі лінії – тренувальні дані, а сині — перевірні. Пунктирні лінії – фактичні значення, суцільні лінії – згладжені значення.

Fig. 2. Graphs of the dependence of accuracy and loss on the number of training epochs (first dataset): accuracy of the SGD optimizer (a), losses of the SGD optimizer (б), accuracy of the Adam optimizer (в), losses of the Adam optimizer (г). Brown lines – training data, blue ones – validation data. Dashed lines – actual data, solid ones – smoothed data.

Крім того, час тренування для SGD був трохи коротшим (11,75 хвилин за 16 епох) порівняно з Adam (13,91 хвилин за 19 епох), що також є перевагою. Враховуючи вищу точність і менший час тренування, оптимізатор SGD виглядає кращим вибором для даного завдання. Сумарно отримані результати для влучності та втрат моделей отримані на першому тренувальному та перевірному наборі даних представлено у таблиці 1.

На рис.3 представлено графіки результатів точності і втрат для тренувальних та перевірних даних другого набору для оптимізаторів SGD та Adam. Для оптимізатора SGD (Рис. 3а – точність, Рис. 3б – втрати) темно-сині лінії показують тренувальні дані, а світло-сині — перевірні. Пунктирні лінії представляють фактичні значення, тоді як суцільні лінії показують згладжені значення. Точність тренувальних даних поступово зростає до згладженого значення 0,2479, а фактичне значення становить 0,2663. Перевірна точність також збільшується, досягаючи згладженого значення 0,239 і фактичного значення 0,2498.

Таблиця 1. Кінцеві значення тренування моделі для першого набору даних.
Table 1. Final values of model training using the first dataset.

		Оптимізатор			
		SGD		Adam	
		Фактичні значення	Згладжені значення	Фактичні значення	Згладжені значення
Точність	Тренувальна	0,8343	0,8297	0,8252	0,8218
	Перевірні	0,8117	0,8141	0,805	0,7929
Втрати	Тренувальні	0,4545	0,4694	0,48	0,4901
	Перевірні	0,5498	0,541	0,5479	0,5812
Час(хв)		11,75		13,91	
К-сть циклів навчання		16		19	

Втрати на тренувальних даних зменшуються до згладженого значення 3,7209 і фактичного – 3,59, тоді як перевірні втрати зменшуються до згладженого значення 3,8222 і фактичного – 3,7522. Загальний час тренування становить 7,282 год.

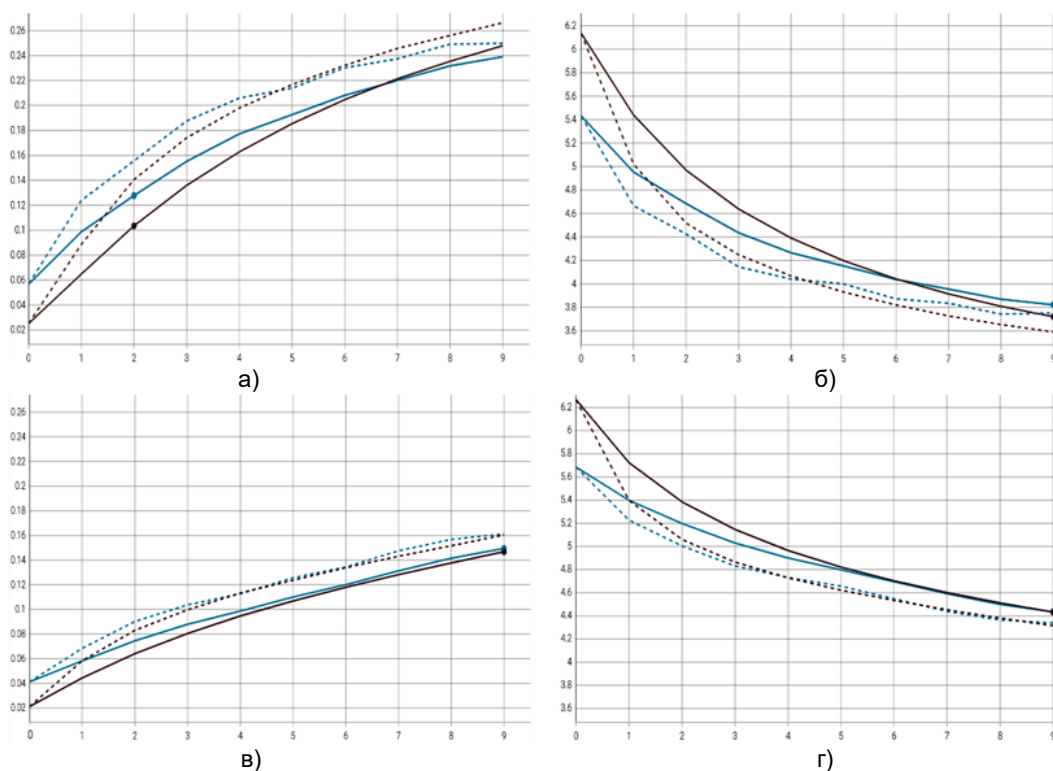


Рис.3. Графіки залежності точності та втрат від кількості епох навчання (другий набір даних):точність оптимізатора SGD (а), втрати оптимізатора SGD (б), точність оптимізатора Adam (в), втрати оптимізатора Adam (г). Коричневі лінії – тренувальні дані, а сині — перевірні. Пунктирні лінії – фактичні значення, суцільні лінії – згладжені значення.

Fig. 3. Graphs of the dependence of accuracy and loss on the number of training epochs (second dataset): accuracy of the SGD optimizer (a), losses of the SGD optimizer (б), accuracy of the Adam optimizer (в), losses of the Adam optimizer (г). Brown lines – training data, blue ones – validation data. Dashed lines – actual data, solid ones – smoothed data.

Для оптимізатора Adam (графіки внизу, зліва – точність, справа - втрати) тем-но-сині лінії також представляють тренувальні дані, а світло-сині — перевірні. Пунктирні лінії відображають фактичні значення, тоді як суцільні лінії показують згладжені значення. Точність тренувальних даних збільшується до згладженого значення 0,1468, а фактичне значення становить 0,1604. Перевірна точність також збільшується до згладженого значення 0,1495 і фактичного значення 0,1612. Втрати на тренувальних даних зменшуються до згладженого значення 4,4333 і фактичного значення 4,3156, тоді як перевірні втрати зменшуються до згладженого значення 4,4347 і фактичного значення 4,3387. Загальний час тренування становить 7,396 годин. Сумарно отримані результати для влучності та втрат моделей отримані на другому тренувальному та перевірному наборі даних представлені у таблиці 2.

Таблиця 2. Кінцеві значення тренування моделі використовуючи другий набір даних.

Table 2. Final values of model training using the second dataset.

		Оптимізатор			
		SGD		Adam	
		Фактичні значення	Згладжені значення	Фактичні значення	Згладжені значення
Точність	Тренувальна	0,2663	0,2479	0,1604	0,1468
	Перевірна	0,2498	0,239	0,1612	0,1495
Втрати	Тренувальні	3,59	3,7209	4,3156	4,4333
	Перевірні	3,7522	3,8222	4,3387	4,4347
Час(год)		7,28		7,4	
К-сть циклів навчання		9		9	

Виходячи з цих результатів, видно, що обидва оптимізатори показують зростання точності та зменшення втрат із збільшенням кількості епох. Однак, SGD досяг кращої точності як на тренувальних, так і на перевірних даних у порівнянні з Adam. Також втрати для SGD були нижчими, що вказує на ефективніше навчання моделі з використанням цього оптимізатора.

На основі даних можна визначити деякі характеристики результатів моделі навчання:

- Зростання точності з кожною епохою навчання, наприклад, від 0,02 до 0,16 з першої до десятої епохи навчання. Це свідчить про те, що модель поступово вдосконалюється під час навчання.
- Стабільність точності на етапі перевірки, наприклад, від 0,04 до 0,161 з першої до десятої епохи перевірки. Це може свідчити про те, що модель добре узгоджується з новими даними після навчання.
- Точність навчання на початку зазвичай невелика, але з кожною новою епохою вона зростає. Наприклад, у першій епосі точність для навчання та перевірки становить 0,02 та 0,04, відповідно, а у десятій епохі ці величини зростають до 0,16 та 0,161, відповідно.
- Потенційне перенавчання спостерігається, якщо точність на навчальному наборі даних значно перевищує точність на перевірному наборі. Однак, у

представленому випадку точності на обох наборах досить близькі, що може свідчити про відсутність перенавчання або його незначний вплив.

Загалом, результати показують, що модель покращується з кожною новою епохою навчання, і точність на перевірки підтверджує стабільність та ефективність моделі.

Тестування роботи моделі для класифікації зображень було проведено для різноманітних зображень. Тестування виконувалось на нейромережі з використанням оптимізатора SGD, який забезпечує кращі результати порівняно з оптимізатором Adam. На рис.4а показано зображення автомобіля швидкої допомоги (з набору даних для навчання (рис.1)), та результат його розпізнавання у форматі JSON (запис об'єктів JavaScript).



(a)



(б)

Рис.4. Тестові зображення та визначені категорії: класифікація зображення швидкої допомоги (а), класифікація зображення легкового автомобіля (б).

Fig. 4. Test images and defined categories: ambulance image classification (a), car image classification (б).

Із рис. 4а видно, що першою визначеною категорією є автомобіль швидкої допомоги (ambulance), тобто категорія визначена цілком вірно. Також, оскільки в автомобіля швидкої допомоги є певні елементи схожості з поліцейським мікроавтобусом, то одразу наступною йде ця категорія. Нижче на рис.4б показано вибране для тесту зображення автомобіля, яке має складний ракурс, тому нейронна мережа визначила, що це зображення позашляховика, хоча і наступна категорія визначена правильно.

ВИСНОВОК

У результаті виконаної роботи було створено систему автоматичної класифікації зображень на основі згорткової нейронної мережі з використанням бібліотеки TensorFlow та системи візуалізації TensorBoard для збереження метрик навчання. Порівнювалися два популярні оптимізатори: Adam та SGD.

Виходячи з аналізу таблиць результатів, оптимізатор SGD показав кращу продуктивність у порівнянні з Adam. Початкова точність тренування та перевірки для SGD була трохи вищою, а швидкість зростання точності суттєво перевершувала

Adam, досягаючи значно вищих значень на 10-й епосі. Остаточна точність моделі з оптимізатором SGD виявилася вищою як на тренувальному, так і на перевірному наборі даних, що свідчить про більш ефективне навчання. Крім того, SGD демонстрував швидше зменшення втрат, знову ж таки забезпечуючи нижчі кінцеві значення втрат порівняно з Adam. Додатково слід зауважити, що час тренування для SGD був трохи меншим, що також є його перевагою порівняно з оптимізатором Adam. Для першого набору даних час тренування для SGD був 11,75 хвилин за 16 епох порівняно з Adam – 13,91 хвилин за 19 епох. В обох випадках моделі не демонстрували значної різниці між тренувальними та перевірними наборами, що свідчить про відсутність перенавчання та добру здатність до узагальнення. Таким чином, для даних моделей та наборів даних оптимізатор SGD був визначений як більш ефективний, забезпечуючи кращу продуктивність та стабільність роботи системи автоматичної класифікації зображень.

ЛІТЕРАТУРА

- [1] Y.Zhang *Why Transformers Need Adam: A Hessian Perspective*/ C.Chen, T.Ding, Z.Li, R.Sun, Z.Luo// Cornell Univer. Machine Learning arXivLabs: [arXiv:2402.16788](https://arxiv.org/abs/2402.16788)
- [2] N. Landro *Mixing ADAM and SGD: a Combined Optimization Method*/ I.Gallo, R.L.Grassa// Cornell Univer. Machine Learning arXivLabs: [arXiv:2011.08042](https://arxiv.org/abs/2011.08042)
- [3] Diederik P. Adam: A Method for Stochastic Optimization/ Kingma, Jimmy Ba// Cornell Univer. Machine Learning arXivLabs: [arXiv.1412.6980](https://arxiv.org/abs/1412.6980)
- [4] P. J. Kamali *Stochastic Gradient Descent outperforms Gradient Descent in recovering a high-dimensional signal in a glassy energy landscape*/ P. Urbani// Cornell Univer. Machine Learning arXivLabs: [arXiv.2309.04788](https://arxiv.org/abs/2309.04788)

DETERMINATION OF THE BEST OPTIMIZER FOR A NEURONETWORK IN THE DEVELOPMENT OF AUTOMATIC IMAGE TAGGING SYSTEMS

Andrian Kozynets 

*Ivan Franko National University of Lviv
50 Dragomanov St., Lviv 79005, Ukraine*

ABSTRACT

Background. Choosing the best optimizer is an important step in developing efficient automatic image classification systems. In particular, for neural networks based on convolutional neural networks (CNNs), the choice between popular optimization methods such as Adam (Adaptive Moment Estimation) and SGD (Stochastic Gradient Descent, SGD) can significantly affect their performance. Comparing these optimizers allows us to determine the most suitable optimizer for solving specific machine learning problems.

Methods and tools. The TensorFlow library and the TensorBoard machine learning visualization interface were used to train and test the models. The neural networks were trained on two different datasets with significantly different characteristics. For each dataset, training was performed using two optimizers: Adam and SGD. When analyzing the results, such parameters as training and validation accuracy, as well as training and validation losses, were taken into account.

Results and Discussion. According to the results, the SGD optimizer showed better performance indicators compared to Adam. The initial accuracy for SGD was higher, and a significant excess of the accuracy growth rate was observed. In addition, the final accuracy

of the model with the SGD optimizer was higher on both the training and validation datasets, which indicates a more efficient training process using this optimizer.

Conclusions. It was determined that the use of the SGD optimizer leads to better results compared to the Adam optimizer for automatic image classification tasks. This is confirmed by higher accuracy indicators and a faster training process.

Keywords: TensorFlow, Convolutional Neural Network, Adaptive Moment Estimation, Stochastic Gradient Descent, GPU, TPU.

УДК 004.03, 004.9

ЕФЕКТИВНІСТЬ ОБРОБКИ ПАРАЛЕЛЬНИХ ЗАПИТІВ У СЕРЕДОВИЩІ AMAZON WEB SERVICES З ВИКОРИСТАННЯМ КОНТЕЙНЕРІВ

Олег Сігунов^{ORCID}, Лідія Демків^{ORCID}

Львівський національний університет імені Івана Франка,
вул. Драгоманова 50, 79005 м. Львів, Україна

Сігунов, О., Демків, Л. (2025). Ефективність обробки паралельних запитів у середовищі Amazon Web Services з використанням контейнерів. *Електроніка та інформаційні технології*, 29, 47–56. <https://doi.org/10.30970/eli.29.5>

АНОТАЦІЯ

Передумови. У статті досліджено ефективність роботи web-застосунку, розгорнутого у хмарних сервісах Amazon веб сервіс (AWS) із використанням контейнеризації на базі AWS сервісу еластичних контейнерів (ECS) та AWS еластичного сервісу Kubernetes (EKS). Особливу увагу приділено аналізу продуктивності системи в умовах високого навантаження, характерного для онлайн-сервісів, що забезпечують відображення спортивних змагань у реальному часі. Оцінено здатність застосунку обробляти велику кількість запитів від користувачів одночасно, що є критично важливим для забезпечення коректної роботи у режимі реального часу.

Методи та інструменти. Методологія дослідження базується на навантажувальному тестуванні із застосуванням сучасних інструментів для генерації запитів та збору метрик. Проведено серію експериментів з моделювання різних сценаріїв використання, включаючи пікові навантаження, стабільний робочий режим та ситуації з різкою зміною трафіку. Аналізувалися ключові параметри продуктивності: швидкість обробки запитів сервером, рівень стабільності системи при зростанні навантаження, а також частка помилково оброблених запитів.

Результати та обговорення. Результати експериментів продемонстрували, що AWS EKS забезпечує найкраще співвідношення продуктивності, стабільності та вартості для розгортання подібних рішень. Використання Kubernetes-кластерів дозволяє гнучко масштабувати інфраструктуру, адаптуючи її під змінні навантаження без значних втрат у швидкодії. AWS ECS, хоча й забезпечує достатню продуктивність, виявився менш гнучким у порівнянні з AWS EKS, особливо у випадках динамічних змін навантаження.

Висновки. Дослідження підтвердило доцільність використання AWS EKS як основної платформи для контейнеризації веб-застосунку з високим навантаженням. Отримані результати можуть бути корисними для розробників та архітекторів хмарних рішень, які працюють із системами реального часу, що вимагають високої продуктивності, гнучкості та надійності. Окрім того, наведені дані можуть слугувати основою для подальших досліджень у сфері оптимізації хмарних ресурсів та підвищення ефективності роботи веб-сервісів.

Ключові слова: хмарні сервіси, контейнери, AWS ECS, AWS EKS, Gatling, сценарії тестування.



© 2025 Oleg Sigunov & Lidiia Demkiv. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ВСТУП

В сучасних цифрових хмарних середовищах багато проектів працюють у режимі реального часу. Забезпечення стабільної та надійної роботи проектів вимагає ефективного використання ресурсів, які пропонують хмарні сервіси для забезпечення масштабованості та стабільності роботи проектів, а також з метою уникнення непередбачуваних витрат на інфраструктуру та обслуговування. Особливо важливою є проблема забезпечення стабільної роботи проектів у хмарних сервісах при непостійному характері споживання ресурсів з боку користувачів. В хмарних сервісах обчислювальні ресурси та дані переносяться на віддалені сервери. Для забезпечення стабільної роботи виникає потреба у виборі оптимальних параметрів хмарних сервісів та створенні ефективних систем для зберігання, оброблення та аналізу даних. Оцінюючи продуктивність та стабільність, важливо враховувати різноманітні параметри, такі як обсяг доступної пам'яті, пропускна здатність мережі та локальні затримки, особливо у контексті контейнеризації. В [1,2] проведено дослідження оптимізації витрат на AWS через розгляд моделей ціноутворення, стратегій та практичних ідей управління ресурсами. Показано, що ретельне планування та визначення пріоритетів завдань дозволяє ефективно мінімізувати витрати часу та ресурсів. У роботі [3] розглянуто технічні аспекти розгортання, керування та масштабування контейнеризованих програм і застосунків в інфраструктурі AWS за допомогою платформи контейнеризації Docker на віртуальних машинах Amazon EC2, а також у середовищах AWS ECS і AWS EKS. В сучасній розробці та розгортанні програмного забезпечення використання контейнеризації часто залежить від конкретних потреб і масштабів проекту. В [4,5] розглянуто особливості контейнеризації та автоматизованого управління контейнерами для оптимізації вартості розгортання мікросервісів. Оскільки контейнеризація стала широко використовуваним підходом, важливо дослідити її переваги для ефективного управління та масштабування інфраструктури. В статті визначається оптимальний сервіс для контейнеризації шляхом проведення навантажувального тестування веб-застосунку, розгорнутого в хмарному сервісі. Розглядаючи різні хмарні сервіси від різних постачальників, необхідно визначити оптимальне середовище та розглянути можливості розміщення системи в різних хмарних середовищах.

МЕТОДИ ТА ІНСТРУМЕНТИ

Вибір середовища розгортання для тестового веб застосунку, розробленого мовою Java та призначеного для трансляції результатів змагань у реальному часі великій кількості користувачів, потребує врахування таких аспектів, як мережева доступність і зручність масштабування. Оптимальним рішенням є запуск сервісу в хмарному середовищі, що має ряд переваг перед використанням стаціонарного сервера. Для розгортання обрано хмарні сервіси Amazon Web Services, які в порівнянні з Microsoft Azure та Google Cloud Platform (GCP) мають свої плюси та мінуси. До переваг AWS належить наявність програмного інтерфейсу взаємодії для всіх сервісів, що дозволяє керувати хмарними ресурсами автоматизовано через код, знижуючи витрати на підтримку. Серед недоліків – дещо вищі витрати в деяких сценаріях використання та складність розгортання проекту, яка вимагає глибоких знань сервісів або підтримки спеціалізованої команди.

У контексті тестового веб застосунку важливим функціоналом є забезпечення стабільного доступу до даних користувачів та отримання результатів у реальному часі. Тому конфігурація системи має бути оптимізована для швидкого відправлення даних багатьом користувачам у короткі проміжки часу. Після вибору хмарного сервісу проведено тестування конфігурацій, розгорнутих в AWS EKS та AWS ECS, щоб обрати оптимальний сервіс для стабільної роботи веб застосунку.

Мова програмування Scala є новим поколінням віртуальної машини Java (JVM – Java Virtual Machine). Scala використовує сучасну та ефективну модель акторів (в даному контексті термін «актори» позначає універсальні примітиви паралельного виконання), що дозволяє розробникам визначати кожен об'єкт як окрему сутність з властивостями, поведінкою та станом. Це спрощує взаємодію між потоками, забезпечуючи більший контроль над даними та підвищуючи ефективність процесів.

Gatling – це відкрите програмне забезпечення для тестування роботи програми, написане мовою Scala. Воно дозволяє обробляти великий об'єм трафіку на одному комп'ютері. Під час тестування досліджується, як система справляється з різними навантаженнями. Багато програм добре працюють тільки при невеликому числі користувачів, але при стрімкому збільшенні кількості користувачів можуть виникати проблеми з роботою системи. Тестування дозволяє виявити проблемні місця та виправити їх для підвищення продуктивності.

Існують кілька типів тестування:

- тестування навантаженням (Load Testing): тестування системи при плановій кількості користувачів та/або трафіку;
- стрес-тестування (Stress Testing): тестування системи за постійного збільшення навантаження для виявлення точки відмови;
- тестування стабільності (Soak Testing): тестування зі стабільним навантаженням протягом тривалого часу.

Gatling має різні сценарії тестування, але для нашої задачі найкраще підходить стратегія `constantConcurrentUsers` яка підтримує певну кількість користувачів (симулює користувачів, які спостерігають за змаганнями).

Основні параметри оцінки продуктивності системи включають час відгуку транзакції, пропускну здатність, відсоток успішних запитів. Gatling надає інформативні звіти для полегшення аналізу результатів.

Amazon ECS дозволяє керувати контейнерами у хмарному середовищі AWS. Amazon EKS – це керована служба кластерів Kubernetes від AWS, яка дозволяє легко запускати, керувати і масштабувати контейнеризовані додатки в хмарному середовищі.

ECS та EKS підтримують два типи розгортання: використання віртуальних машин у хмарному середовищі (інстанси) та повністю автоматизований сервіс Fargate, який керує вибором, масштабуванням та управлінням інстансами. Для тестування web-застосунку використано варіант EC2 для того, щоб моделювання відбувалось за однакових ресурсів.

AWS Elastic Load Balancer (ELB) розподіляє вхідний трафік між різними компонентами системи, підвищуючи її надійність та масштабованість. Amazon Elastic Container Registry (ECR) дозволяє керувати реєстрацією контейнерів у хмарному середовищі AWS, забезпечуючи зберігання, управління та розповсюдження контейнерів. AWS Relational Database Service (RDS) надає можливість розгортання реляційних баз даних у хмарному середовищі AWS, включаючи вибір типу бази даних та її двигуна для оптимальної продуктивності та надійності.

Web-застосунок написано за допомогою Java програмного каркасу Spring, а проект для тестування використовує Scala. Під час розробки використано конструктор залежностей (Maven) та середовище розробки IntelliJ IDEA. Програма складається з двох незалежних між собою мікросервісів: один з них відповідає за реєстрацію та менеджмент користувачів, а інший за передачу даних в реальному часі. Маршрутизація вхідних запитів та спілкування між сервісами налагоджено в межах сервісів. Для забезпечення максимальної близькості до реальних умов, застосунок було розгорнуто як ізольоване середовище - докер контейнер, що містить застосунок разом із усіма необхідними залежностями (бібліотеками, конфігураціями тощо)

Використання контейнеризації має ряд переваг. По-перше, контейнеризація забезпечує консистентність середовища розробки, тестування і виробництва, що

дозволяє уникнути проблем, пов'язаних з різницею між конфігураціями. Крім того, контейнери дозволяють ефективно використовувати ресурси сервера, оскільки вони можуть бути миттєво розгорнуті та знищені за потреби. Друга важлива перевага полягає в легкості перенесення додатків між різними середовищами, що спрощує розгортання та масштабування додатків. Контейнеризація також підвищує безпеку, оскільки вона ізолює додатки один від одного і дозволяє керувати доступом до ресурсів. Використання контейнеризації сприяє швидкому розгортанню нових версій додатків і має позитивний вплив на розробку та впровадження нових функцій шляхом зменшення часу, потрібного для тестування та перевірки змін.

Зображена на рис.1 процедура розгортання застосунку для тестування у хмарному середовищі включає декілька кроків. Для застосунку створеного з використанням Spring Boot Application здійснюється побудова Docker-образу, який може бути завантажений до реєстру (наприклад, Docker Hub або AWS ECR). Після завантаження образу до реєстру для нього буде згенеровано публічне посилання, яке необхідно вказати під час налаштування конфігурацій кластерів, таких як Amazon ECS або Amazon EKS. Це дозволить кластерам коректно використовувати цей образ під час запуску контейнерів, оскільки один і той самий образ може мати декілька версій від кількох різних дистриб'юторів.

Архітектура системи для тестування, розгорнутої за допомогою ECS/EKS, зображена на схемі рис. 2. Основні компоненти включають Application Load Balancer (ALB), який розподіляє запити між контейнерами в кластері. Кластер складається з набору інстансів, на яких розгорнуті сервіси (Spring Boot application), створені з образів за допомогою сервісу Elastic Container Registry (ECR) для зберігання, управління та розповсюдження контейнерних образів Docker у хмарі AWS, використаних для конфігурації кластера. Також система включає базу даних завдяки використанню керованого сервісу реляційної бази даних від AWS, що зберігає дані, які використовуються застосунком. Ця архітектура забезпечує надійну та масштабовану інфраструктуру для тестування додатків, що дозволяє ефективно розподіляти навантаження та працювати з необхідними даними.

Необхідно порівняти продуктивність застосунку, який запущений в ECS або EKS при умові рівних ресурсів (однаковий тип інстансу, однаковий програмний код) та визначити оптимальну конфігурацію інстансів, яка буде одночасно витримувати велику кількість користувачів. В Amazon ECS та Amazon EKS можна використовувати

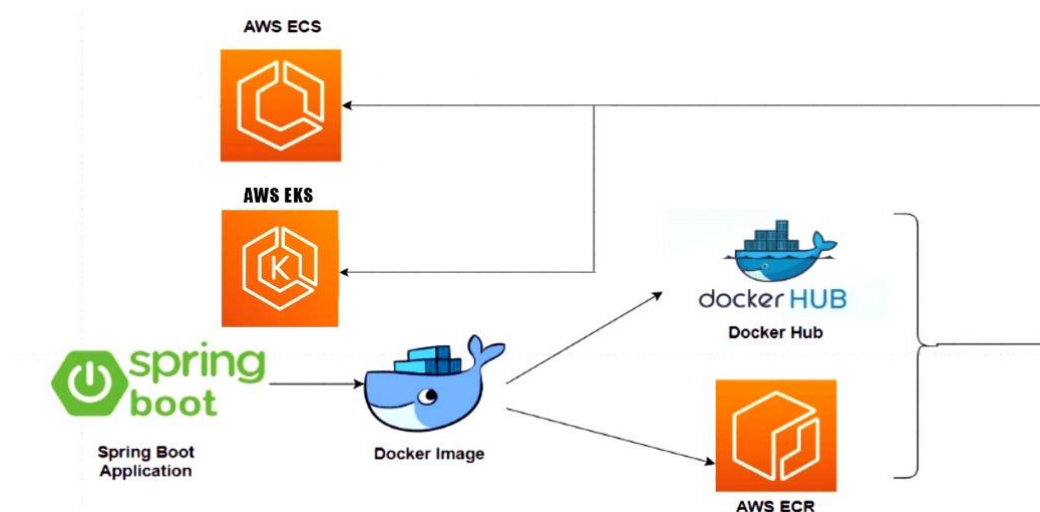


Рис. 1. Схема розгортання web-застосунку для тестування за допомогою ECS/EKS.

Fig. 1. Deployment scheme of the web application for testing using ECS/EKS.

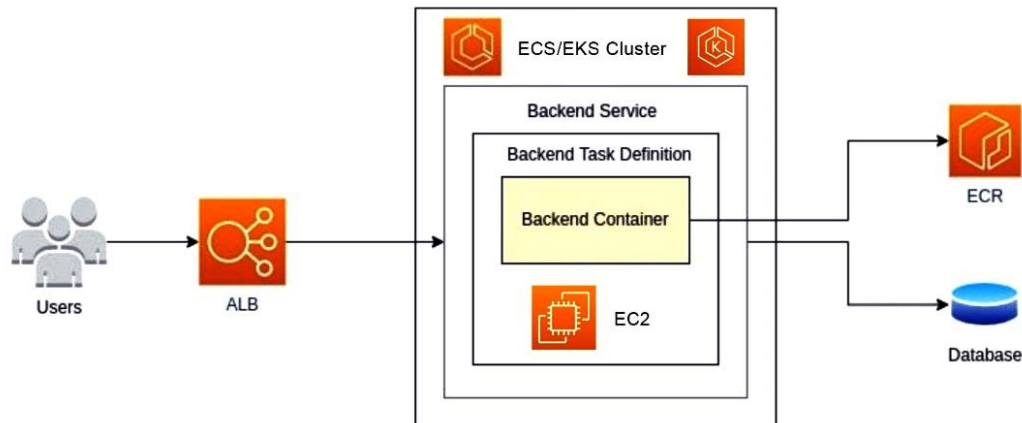


Рис. 2. Архітектура системи для тестування розгорнутої за допомогою ECS/EKS.
Fig. 2. System architecture for testing deployed using ECS/EKS.

EC2 сервери, які представлені багатьма типами з різними ключовими характеристиками. Сервери призначені для вирішення різноманітних задач у різному ціновому діапазоні. Вони дозволяють обрати найбільш оптимальні ресурси для конкретних потреб. Для дослідження використано інстанси t2.small, які належать до групи загального застосування. Інстанси забезпечують баланс між продуктивністю та вартістю, мають однопоточний віртуальний центральний процесор (1 vCPU) та 2 ГБ оперативної пам'яті, що робить їх ідеальним вибором для широкого спектру завдань, включаючи розробку, тестування від невеликих до середніх систем в продуктивних середовищах.

MySQL база даних web-застосунку розміщена у хмарному сервісі за допомогою AWS RDS. Для бази даних обрано сервер db.r5.large, який базується на 2vCPU 16 ОЗУ та 5 гбіт/с мережі, що для задач web-застосунку дуже надлишково, але це необхідно для того, щоб впевнитись, що тестування перевіряє саме сервери програми, а не базу даних.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

Для тестування розроблено сценарій, який є наближеним до реальної моделі роботи застосунку. Користувачі відправляють запит на сервер для отримання даних турнірних таблиць у реальному часі. Кількість одночасних користувачів варіювалась для оцінки впливу навантаження на систему. Тести проведені для 20, 30, 50 та 80 користувачів, які одночасно надсилають запит на сервер за принципом повторного надсилання запиту після отримання відповіді на попередній. Цей підхід дозволив перевірити не лише працездатність системи в умовах реального завантаження, але й з'ясувати, як вона поводить себе за різних умов, що важливо для оцінки її ефективності та надійності. Тривалість кожного проведеного тестування дорівнювала 5 хв. Така тривалість дає змогу показати як поводить себе застосунок під постійним навантаженням, оскільки цього часу достатньо, щоб виявити вузькі місця системи. Такі тести є об'єктивними і для триваліших проміжків часу.

Зведені результати тестування для всіх рівнів навантаження представлено у таблицях 1 та 2 (для ECS та EKS, відповідно). У лівій колонці таблиці представлено параметри: мінімальний, максимальний та середній час відповідей сервера у мілісекундах, кількість успішних чи невдалих відповідей сервера.

З даних таблиць бачимо, що однаковою для EKS та ECS є те, що зі зростанням кількості користувачів росте також середній час відповіді, при чому росте майже лінійно. З результатів тестування для EKS випливає, що середній час обробки запиту

Таблиця 1. Результати тестування для ECS
Table 1. Testing results for ECS

	Кількість одночасних користувачів			
	20	30	50	80
Мінімальний час відповіді, мс	51	51	52	51
Максимальний час відповіді, мс	20033	20030	20030	20037
Середній час відповіді, мс	1260	1876	3470	4365
Кількість успішних запитів	4579	4580	4146	4655
Кількість не успішних запитів	186	237	185	872
Відсоток не успішних запитів	4%	5%	4%	16%

Таблиця 2. Результати тестування для EKS
Table 2. Testing results for EKS.

	Кількість одночасних користувачів			
	20	30	50	80
Мінімальний час відповіді, мс	51	51	51	51
Максимальний час відповіді, мс	3103	67455	81087	87388
Середній час відповіді, мс	79	84	187	225
Кількість успішних запитів	72574	74633	77066	90045
Кількість не успішних запитів	0	0	5	42
Відсоток не успішних запитів	0%	0%	0%	0%

значно менший (<100 ms), ніж для ECS. Для ECS час обробки запиту завжди більше 1 секунди, через що кількість опрацьованих запитів за однаковий час відрізняється дуже суттєво: більше 70000 тисяч для EKS та 5 тисяч для ECS.

Цікавим спостереженням є те, що деякі запити для EKS займали понад хвилину часу, однак все ще були успішно виконаними. В той же час для ECS не має запитів, які тривали більше 20 секунд, однак кількість не виконаних запитів значно більша. Ці результати пов'язані з тим, як система Kubernetes керує вхідними запитами. В Kubernetes запити можуть стояти в черзі довше, однак зрештою вони будуть опрацьовані. На всіх рівнях тестування в EKS менше 1% похибок для всіх рівнів навантаження, на відміну від ECS, де серед розглянутих на максимальному рівні запитів - 16% не є успішними.

Gatling генерує звіти в графічній формі. Нижче наведені звіти з тестування ECS та EKS при однакових навантаженнях у 80 користувачів на секунду.

З діаграм на рис.3 видно, що в обох випадках найбільша частина запитів оброблялася швидше за 800 мс. Однак, для ECS суттєва частка запитів оброблялася в діапазонах 800-1200 мс та понад 1200 мс. Також велика кількість запитів є неуспішною. В той час, як для EKS, дуже мала частка запитів оброблялася в часовому діапазоні 800-1200 мс, а кількість невдалих запитів є несуттєвою (менше 0,5%).

З графіків на рис.4 видно, що система на ECS більшу частину часу підтримувала стабільну кількість користувачів, близько 150. Проте, оскільки це є фактичним максимумом для цієї системи, час відповіді на запити змінювався. На графіку видно

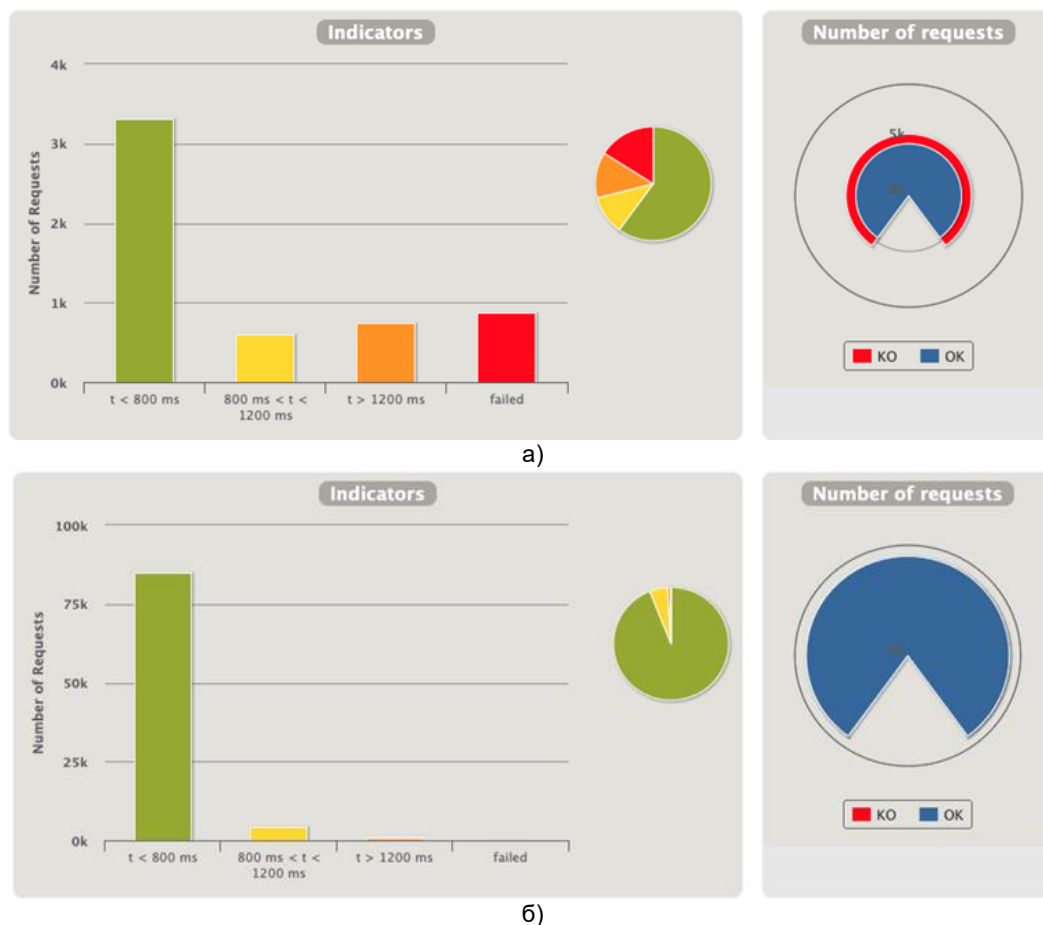


Рис. 3. Діаграми розподілу часу відповіді на запит для а) ECS та б) EKS.
Fig. 3. Response time distribution diagrams for (a) ECS and (b) EKS.

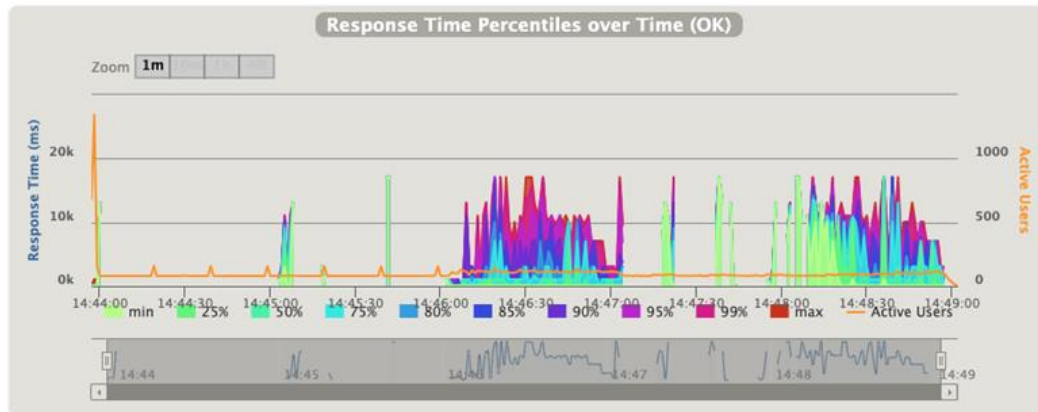
кілька стрибків, які показують періоди, коли запити оброблялися близько 20 секунд і відповідно потрапляли у 90-й перцентиль.

У той же час система на EKS приймала значно більшу кількість запитів і, відповідно, більшу кількість користувачів з деякими піками понад 1000 користувачів. Попри це, вона забезпечувала дуже стабільний час відповіді, який потрапляв у межі 50-го перцентиль. Це свідчить про високу ефективність і продуктивність системи на EKS у порівнянні з ECS.

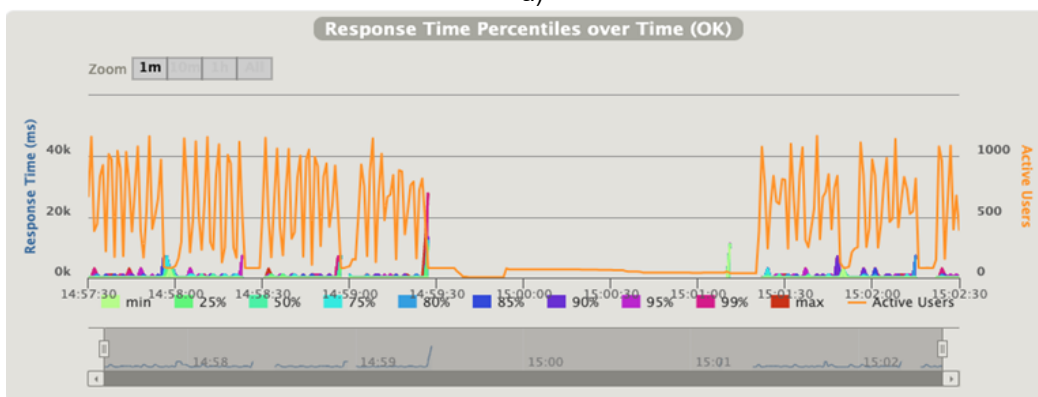
Детальний аналіз показує, що ECS, при навантаженні близько 150 користувачів, досягала свого максимального потенціалу, що призводило до значних затримок у відповіді на запити. Це особливо помітно під час пікових навантажень, коли час відповіді перевищував 20 секунд.

На відміну від ECS, система на EKS демонструвала високу стійкість і ефективність при значно більшому навантаженні. Навіть при пікових навантаженнях коли в системі було понад 1000 користувачів, EKS продовжувала забезпечувати стабільний і швидкий час відповіді. Це підкреслює її здатність обробляти великі обсяги запитів без суттєвих затримок, що є важливим показником продуктивності та надійності системи.

З даних таблиць видно, що кількість подій за секунду, які обробляє ECS, значно менша за кількість подій, які може обробити EKS. Зокрема, ECS обробляє лише 18 подій на секунду, тоді як EKS обробляє 299 подій на секунду.



а)



б)

Рис. 4. Графік відношення кількості активних користувачів до часу відповіді для а) ECS та б) EKS
Fig. 4. Graph of the relationship between the number of active users and response time for (a) ECS and (b) EKS.

Час відповідей для 50-го та 75-го процентилів зростає дуже незначно для EKS (73 мс та 82 мс відповідно). Проте для ECS цей розрив є величезним: 63 мс для 50-го перцентиліа та 3203 мс для 75-го перцентиліа. Також 95- й та 99-й перцентилі для ECS майже не відрізняються і дуже близькі до максимального часу відповіді. Це свідчить про те, що система на ECS вже працює за межами своєї продуктивності, а кількість помилок перевищує допустиму норму. В той самий час, 95-й та 99-й перцентилі для EKS все ще плавно зростають і знаходяться досить далеко від максимального часу відповіді. З рис.5 видно відмінності в типах помилок для обох систем. У випадку з ECS через надмірну кількість запитів система перестає їх приймати, що призводить до виникнення помилки `ConnectionTimeoutException`. У випадку з EKS помилка має дещо інший характер. Вона свідчить про те, що сервіс не встиг віддати відповідь протягом однієї хвилини, через що отримуємо помилку `RequestTimeoutException`.

ВИСНОВОК

Створено веб-застосунок, а також протестовано його швидкодiю та стабiльнiсть роботи з урахуванням рiзних засобiв для контейнеризацiї за допомогою AWS ECS та AWS EKS. Виявлено, що найбільшою рiзницею мiж способами контейнеризацiї є механiзм обробки вхiдних запитiв. З порiвняння видно, що система на ECS показала себе значно стабiльнiшою у пiдтримцi однакової кiлькостi користувачiв без значних

STATISTICS Expand all groups | Collapse all groups

Requests ^	Executions					Response Time (ms)							
	Total	OK	KO	% KO	Cnt/s	Min	50th pct	75th pct	95th pct	99th pct	Max	Mean	Std Dev
Global Information	5527	4655	872	16%	18.062	51	63	3203	20022	20030	20037	4365	7421
Get Info	5527	4655	872	16%	18.062	51	63	3203	20022	20030	20037	4365	7421

ERRORS

Error	Count	Percentage
l.n.c.ConnectTimeoutException: connection timed out: ecs3-2009852195.eu-central-1.elb.amazonaws.com/3.64.203.82:8080	442	50.688 %
l.n.c.ConnectTimeoutException: connection timed out: ecs3-2009852195.eu-central-1.elb.amazonaws.com/52.58.41.221:8080	430	49.312 %

a)

STATISTICS Expand all groups | Collapse all groups

Requests ^	Executions					Response Time (ms)							
	Total	OK	KO	% KO	Cnt/s	Min	50th pct	75th pct	95th pct	99th pct	Max	Mean	Std Dev
Global Information	90087	90045	42	0%	299.292	51	73	82	1063	1131	87388	225	2030
Get Info	90087	90045	42	0%	299.292	51	73	82	1063	1131	87388	225	2030

ERRORS

Error	Count	Percentage
l.g.h.c.l.RequestTimeoutException: Request timeout to a032ce26345cb4ea5b299d26356bb7d5-184144236.eu-central-1.elb.amazonaws.com/3.67.200.26:80 after 60000 ms	24	57.143 %
l.g.h.c.l.RequestTimeoutException: Request timeout to a032ce26345cb4ea5b299d26356bb7d5-184144236.eu-central-1.elb.amazonaws.com/3.66.208.62:80 after 60000 ms	13	30.952 %
l.g.h.c.l.RequestTimeoutException: Request timeout after 60000 ms	5	11.905 %

б)

Рис. 5. Детальні результати тестування для а) ECS та б) EKS

Fig. 5. Detailed testing results for (a) ECS and (b) EKS.

стрибків протягом роботи. ECS може мати перевагу в обробці запитів, які вимагають тривалих обчислень. Однак, EKS, хоч і обробляє запити значно швидше, однак приймає їх хвилями, що приводить до значних стрибків у кількості активних користувачів. Водночас, кількість неуспішних запитів, які оброблялися значно довше за середній час на стороні EKS, була значно меншою порівняно з ECS. При обробці багатьох одночасних запитів за допомогою ECS відбувається сповільнення обробки всіх запитів на певний час. Це є критичним недоліком для розглядуваного класу задач, зокрема для відображення даних про змагання. Таким чином, з порівняння видно, що при використанні однакових ресурсів, EKS підходить значно краще для обробки великої кількості простих запитів. Після проведених тестів визначено, що найкраще для реалізації даного проекту підходить AWS EKS.

АВТОРСЬКИЙ ВНЕСОК

Концептуалізація [О.С., Л.Д.]; методика [О.С., Л.Д.]; дослідження [О.С., Л.Д.]; написання початкової версії статті [О.С.]; перегляд та редагування [Л.Д.]; підготовка графічного матеріалу [О.С.]; загальне керівництво [Л.Д.].

Усі автори ознайомилися та погодилися з опублікованою версією статті.

ЛІТЕРАТУРА

- [1] Bhupesh Kumar Dewangan S, Amit Agarwal, Tanupriya Choudhury, Ashutosh Pasricha Cloud Resource Optimization System Based on Time and Cost // International Journal of Mathematical, Engineering and Management Sciences Vol. 5, No. 4, 758-768, 2020. <https://doi.org/10.33889/IJMEMS.2020.5.4.060>

- [2] Sumanth Tatineni Cost Optimization Strategies For Navigating The Economics Of Aws Cloud Services //International journal of advanced research in engineering & technology, 10(6):827-842, 2019.
https://iaeme.com/Home/article_id/IJARET_10_06_084
- [3] Shimon Ifrah Deploy Containers on AWS: With EC2, ECS, and EKS // ISBN: 978-1-4842-5100-3, 2019. <https://doi.org/10.1007/978-1-4842-5101-0>
- [4] Li, W., Li, X., Chen, L. Pattern learning for scheduling microservice workflow to cloud containers. // Int. J. Mach. Learn. & Cyber., 2024. <https://doi.org/10.1007/s13042-024-02115-5>
- [5] Dasith Edirisinghe, Kavinda Rajapakse, Pasindu Abeysinghe, Sunimal Rathnayake Cost-Optimal Microservices Deployment with Cluster Autoscaling and Spot Pricing // arXiv:2405.12311. 2024. <https://doi.org/10.48550/arXiv.2405.12311>

RESEARCH ON THE EFFICIENCY OF PROCESSING PARALLEL REQUESTS BY AMAZON WEB SERVICES USING CONTAINERS

Oleg Sigunov, Lidiia Demkiv

Ivan Franko National University of Lviv, 50 Dragomanov St., Lviv 79005, Ukraine

ABSTRACT

Background. The article explores the efficiency of a web application deployed in AWS cloud services using containerization based on AWS ECS and AWS EKS. Particular attention is paid to analyzing system performance under high load conditions typical for online services that provide real-time sports competition updates. The ability of the application to handle a large number of user requests simultaneously is assessed, which is critically important for ensuring proper operation in real-time mode.

Methods and tools. The research methodology is based on load testing using modern tools for generating requests and collecting metrics. A series of experiments were conducted to model different usage scenarios, including peak loads, stable operating conditions, and situations with sudden traffic changes. Key performance parameters were analyzed: server request processing speed, system stability under increasing load, and the proportion of unsuccessfully processed requests.

Results and discussion. The experimental results demonstrated that AWS EKS provides the best ratio of performance, stability, and cost for deploying similar solutions. The use of Kubernetes clusters allows for flexible infrastructure scaling, adapting to variable loads without significant losses in speed. While AWS ECS ensures sufficient performance, it proved to be less flexible compared to AWS EKS, particularly in cases of dynamic load changes.

Conclusions. Thus, the study confirmed the feasibility of using AWS EKS as the primary platform for containerizing a high-load web application. The obtained results may be useful for developers and cloud solution architects working with real-time systems requiring high performance, flexibility, and reliability. Additionally, the presented data may serve as a basis for further research in optimizing cloud resources and improving the efficiency of web services.

Keywords: cloud services, containers, instances, AWS ECS, AWS EKS, Gatling, testing scenarios.

UDC 004.93'1, 519.684

OPTIMIZATIONS OF DEEP LEARNING OBJECTS DETECTION MODELS FOR INFERENCE ACCELERATION ON GENERAL-PURPOSE AND HARDWARE-ACCELERATED SINGLE-BOARD PLATFORMS

Dmytro Myroniuk^{ID}, **Bohdan Blagitko**^{ID}
Ivan Franko National University of Lviv,
107 Tarnavsky St., UA–79017 Lviv, Ukraine

Myroniuk, D. M., Blagitko, B. Ya. (2025). Deep Learning objects detection through model optimization to accelerate the process on general-purpose single-board platforms. *Electronics and Information Technologies*, 29, 57-68. <https://doi.org/10.30970/eli.29.6>

ABSTRACT

Background. Description and preparation of modern approaches for deep learning object detection models are provided. Deep learning frameworks for model training and inference, such as TensorFlow and TensorFlow Lite, are used as bases. The concepts of deep learning model optimization are analyzed.

Materials and Methods. The quantized int8 models are used as a baseline for optimization effectiveness estimation. The delegation approach includes software or hardware-optimized variants of neural operations. It prepared to speed up the inference process on target devices. The device with reduced performance resources or microcontroller without floating-point blocks uses a case of base-optimization model with int8 weights. The TensorFlow Lite framework has various quantization types outlined in a detailed explanation. Benchmarks for modern single-board devices are ready, and the correlation between using different optimization approaches, types of single-board platforms, and model inference speed analyses.

Results and Discussion. All tested models are pretrained using the MS COCO dataset (80 classes). All models were prepared for the experiment with 8-bit full integer quantization and output-TF Lite model generation using TensorFlow Object Detection API Docker images and Python 3.11. The testing data samples are obtained from the MS COCO validation dataset archive. The size of the image input is 640x640 RGB. The comparison of image recognition time to 640x640 RGB was conducted on Raspberry Pi 5, Raspberry Pi 4, and Jetson Nano 2GB. Only the Raspberry Pi 5 target device achieved real-time execution (100 ms at most or one fps) as it has more CPU performance than other devices.

Conclusion. Confirmation of the real-time execution approach was achieved by using reference models with reduced image sizes (320x320 RGB). TensorFlow standard model Zoo models, compiled with the TensorRT compiler, were used for the Jetson Nano target as an NPU-optimized case. Real-time execution (100 ms at most or one fps) is reaching for most models and target devices. Such an approach is suitable for less powerful devices with ARM Cortex-A processors.

Keywords: single-board computers, modeling, benchmarking, neural networks, object detection, optimization.

INTRODUCTION

The deep learning model inference of low-power devices has become main stream in several years. That kind of device class combines optimal performance. It is suitable for



© 2025 Dmytro Myroniuk & Bohdan Blagitko. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

running deep-learning models with low power consumption. The model's optimization is one of the most dominant parts of modern deep learning topics. The basic trend of last year is using optimized models with low-cost and low-power MCUs and microcomputers. The based on such platforms solutions are more autonomous than systems based on the cloud solutions desktop. It has enough performance to make inferences in real time.

The single-board computer solutions analyzed for use with object detection deep learning algorithms in this article. These models have more computational complexity than image classification or localization tasks. That kind of model needs more powerful devices. The benchmarks for different object detection models are efficient. The correlations between general-purpose and hardware-accelerated platforms analyses, also. All types of platforms used: general-purpose (Raspberry Pi 4 and Raspberry Pi 5 [2]), platforms with internal accelerator (NVIDIA Jetson Nano 2 GB [3] with internal CUDA-based GPU with 128 cores), general-purpose platforms with external HW accelerator (Raspberry Pi 4, 5 with Google Coral Accelerator [4] as TPU) as target platforms.

LITERATURE REVIEW

Model benchmarking.

Model benchmarking is the most essential task for model testing on target devices. Several organizations are preparing benchmarks for the target device's fundamental performance estimation. The most authoritative organization for such a task is MLPerf [5], which has ready base submission standards for HW vendors. In addition, deep learning software vendors prepare base benchmarks for different devices, for example, Tensorflow Object Detection API models archive [6] and Ultralytics YOLOv5-v8 [7]. However, such vendors do not cover all devices, especially modern single-board target devices with external accelerations.

Deep Learning models for object detection task deployment in real-time systems.

Real-time object detection is one of the most popular tasks in the modern object detection sphere. Here are several approaches to real-time object detection with different configurations.

Article [8] presented benchmarking for different types of optimizations for Raspberry Pi devices. Authors have prepared several models with applied optimization techniques. However, this benchmark does not cover the newest models of Raspberry Pi devices (Raspberry Pi 5) with applied hardware accelerations. This research also does not cover modern object detection models, moreover. This is used widely in the target topics.

Real-time applications have become a prevalent task in modern information technologies; the Articles [9] and [11] present applying deep learning object detection models for real-world applied tasks: peach fruit detection and weed control application. In the first case, the authors used a configuration based on a single-board computer Raspberry Pi 4 with 8 GB of RAM and Google Coral accelerator as an external accelerator. In the other case, the authors used the Jetson Nano module and TensorRT inference engine.

The authors described examples of benchmarking YOLOv3 models on edge mobile devices based on Cortex A-series CPUs [10]. The research was conducted on a Samsung Galaxy Note 8. It is also applicable to Cortex A-series processors.

MATERIALS AND METHODS

Used target devices

All types of single-board computers were used in research as target devices. It can be used for practical cases in various categories. It combines good performance with small power consumption of the devices. Configurations of that kind determined some challenges

and limitations in the deep learning inference on target devices. Single-board computers with different hardware configurations are used for deep learning model inference to improve performance in this article:

General purpose devices class – Raspberry Pi 4 and 5 were used for this case; Computers of that kind have become a standard of general-purpose single-board computers in recent years. All devices were configured for mobile deep learning using appropriate frameworks (TensorFlow Lite in our case). In this research, different configurations of Raspberry Pi model B are used:

1) Raspberry Pi 4 model B, 4 GB. Table 1 gives Details of the OS and target configurations.

A general-purpose device with the external accelerator class. Some single-board platforms can accelerate deep learning model inference via internal GPUs and special ARM delegates, with CMSIS-NN library as an SW acceleration. Several external accelerators for general single-board devices can be used to increase overall performance for deep learning model inference. Google Coral Accelerator is used here as a member of such devices. Details about the HW configuration of the Google Coral accelerator are given in Table 3.

2) Raspberry Pi 5 model B 4 GB. Table 2 gives Details of the OS and target configurations.

Table 1. Raspberry Pi 4 model B HW details

Parameter	Value
Model	Raspberry Pi 4 Model B 4 GB
CPU	Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
RAM	4GB LPDDR4-3200 SDRAM
OS	Raspberry Pi OS (Bullseye)

Table 2. Raspberry Pi 5 model B HW details

Parameter	Value
Model	Raspberry Pi 5 Model B 4 GB
CPU	Quad-core 64-bit Arm Cortex-A-76 @ 2.4 GHz
RAM	LPDDR4X-4267 SDRAM 4GB
OS	Raspberry Pi OS (Bullseye)

Table 3. Google Coral Accelerator HW details

Parameter	Value
Model	Google Coral Accelerator
Computational performance	int8:4 TOPS
Supported types	Quantized integer (8-bit)
Power consumption	2 W
Used SW	Google Edge TPU library

Specialized single-board computers with internal accelerators. Such a class of devices were designed for accelerated computations, such as deep learning models running, parallel computations, machine learning, etc. Here is an example of such devices:

- Nvidia Jetson Nano/Orin/Xavier;
- Google Coral Dev Board;
- Hailo AI (as an extension of Raspberry Pi 5 via M2 HAT+ interface).

As a main difference between such devices, we can specialize in internal integration with board interfaces and using high-speed internal bus for connection. Also, some devices have extended support of quantized models: float16 or int16x8 quantization types are supported for NVidia devices, so in some particular cases, we can achieve better results for quantized models. NVidia Jetson Nano 2 GB development board is used as a member of such a category. Details of OS and target configurations are given in Table 4.

Table 4. Jetson Nano 2 GB HW details

Parameter	Value
Model	NVidia Jetson Nano 2 GB
CPU	Quad-core Cortex-A57 64-bit SoC @ 1.43GHz
RAM	2GB LPDDR4 1600 MHz SDRAM
GPU	Nvidia Maxwell architecture with 128 NVidia CUDA Cores, 921 MHz
OS	Ubuntu Tegra OS, based on Ubuntu 18.04

Used software

Deep Learning frameworks

TensorFlow and TensorFlow Lite are a base deep learning framework for model training and inference. This framework has vast functionality for model training, inference, and converting to smaller data types. The TensorFlow Lite sub-framework has its model interpreter, which can support various deep learning models. TensorFlow Lite also supports inference using CPU optimizations, such as XNNPACK and ARM-NN delegate.

As some of the used models for testing use PyTorch – TFLite model conversion as a middle format. The ONNX format is also one of the most used model formats in modern deep learning. It also supports various optimizations, such as delegating using ARM software runtime optimizations (based on CMSIS-NN library) and XNNPACK delegate. Also, the ONNX format supports parallelization between CPU cores to improve inference speed for multi-core processors.

The NVidia computers use TensorRT, an optimized framework and inference engine. The ONNX and a special TensorRT wrapper convert the functionality of the base models to TensorRT format.

Used optimizations

Several types of optimizations are used in this research. The quantized int8 models are used as a baseline for optimization effectiveness estimation. The following model optimizations are namely:

- TensorRT model optimization and inference on its engine;
- ARM NN delegate for CPU operation optimizations;
- PyCoral tools for Google Coral TPU optimization.

Software delegation of several cases used for the base software optimization as default int8x8 models. The delegation approach includes software or hardware-optimized variants of neural operations. It prepared to speed up the inference process on target devices.

All used single-board computers are based on ARM Cortex A-series processors with ARM architectures. It is used for the base software optimization ARM ACL library.

The ARM NN is for Cortex A-series processors by a special inference engine and Mali GPUs maintained by ARM. It provides one of the most optimized software implementations for NN operations. The ARM NN can be included in the TFLite inference engine via a delegation approach for Python or used directly via C++ and CMake.

The second popular software optimization used in this research is XNNPACK. XNNPACK is a highly optimized solution for various CPU architectures, such as x86, ARM, RISC-V, etc. Google developed it. XNNPACK provides low-level software implementations for neural operations. It can be included in the TFLite interpreter as a delegate.

The research uses the PyCoral library as a hardware-optimized delegate. PyCoral is to run the Google Coral accelerator as a tool for optimizing neural operations. It includes several tools for model preparation and a dedicated TFLite delegate. It is used for Google Coral devices as the inference engine.

Used quantization configurations

Models for mobile inference use optimized models in addition to standard models. Most modern acceleration systems for mobile platforms use model quantization, which converts trained float32 models to data types with reduced memory usage representation. Devices with reduced performance resources or microcontrollers without floating-point blocks can use that model. This work uses a case of base-optimization model with int8 weights. Such models will have reduced computational complexity but can also have reduced accuracy parameters. Table 5 describes a detailed explanation of different quantization types for the TensorFlow Lite framework [12].

Table 5. Quantization types benefits

#	Technique	Benefits	Hardware
1	Dynamic range quantization	4x smaller, 2x-3x speedup	CPU
2	Full integer quantization	4x smaller, 3x+ speedup	CPU, Edge TPU, Microcontrollers
3	Float16 quantization	2x smaller, GPU acceleration	CPU, GPU

Used models

The models for this research use popular mobile object detection models. Such models can be deployed to smaller devices as they use lighter architectures than classic object detection models, which are appointed to state GPUs with large numbers of shader cores. So, the most important criterion for such models is less computational complexity. It will make them more suitable for ARM-based processors of target boards.

As the baseline models for this research, models with input size 640x640 RGB are used. Such an input size is the most popular for modern object detection models. It can save most of the valuable features of recognized objects and provide better accuracy than models with smaller sizes.

On the other hand, models with input size 320x320 RGB are used to provide a segmental recognition approach. Such a class of models still can provide suitable accuracy for large objects. It is less accurate for smaller objects as such size cannot provide enough key features to make a confident forecast.

All tested models were pre-trained using the MS COCO dataset [13] (80 classes) and tested using the validation dataset of the MS COCO 2017 competition. Table 6 provides a list of all researched models with appropriate sources.

RESULTS AND DISCUSSION

Baseline estimation

Inference on target devices with full input size models prepared as the baseline experiment. All used models were ready for the experiment with 8-bit full-integer quantization and output .tflite models generation using TensorFlow Object Detection API Docker images and Python 3.11 as the basic programming language on the reference x86 device. Table 7 presents the testing results for this experiment.

Experiment parameters: TF/TFLite, standard TFLite 8-bit full integer quantization, and Python 3.11.9. The testing data samples are obtained from the MS COCO validation dataset archive. The size input image is 640x640 RGB.

Table 6. Used models

Architecture	Source	Input size	Parameters number
SSD-MobileNetV1	TF OD API [6]	640x640	31,200,598
SSD-MobileNetV2	TF OD API [6]	640x640	2,870,772
EfficientDet_Lite3x	TF OD API [6]	640x640	9,872,762
YOLOv5s	Ultralytics YOLOv5 [7]	640x640	7,026,307
SSD-MobileNetV1	TF OD API [6]	320x320	6,861,134
SSD-MobileNetV2	TF OD API [6]	320x320	2,741,320
EfficientDet_Lite0	TF OD API [6]	320x320	3,366,672
YOLOv5n	Ultralytics YOLOv5 [7]	320x320	1,767,283

Table 7. The speed estimation inference for models with input size 640x640 RGB on target devices

Model name	Model optimization	$\bar{x}$, ms	Σ , ms
Raspberry Pi 4 Input size 640x640			
Basic models			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	Basic Tflite model	2467.0	109.0
ssd_mobilenet_v2_fpn_lite_640x640_coco17 [6]	Basic Tflite model	244.5	6.0
efficientdet_lite3x_640_ptq [6]	Basic Tflite model	743.5	156.5
yolov5s [7]	Basic Tflite model	376.0	6.0
SW-optimized			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	ARM NN Delegate	2117.4	42.6
ssd_mobilenet_v2_fpn_lite_640x640_coco17 [6]	ARM NN Delegate	597.0	10.4
efficientdet_lite3x_640_ptq [6]	ARM NN Delegate	1046.0	28.0
yolov5s [7]	ARM NN Delegate	427.3	11.6
HW accelerated			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	Coral	430.94	13.6
ssd_mobilenet_v2_fpn_lite_640x640_coco17 [6]	Coral	372.57	5.93
efficientdet_lite3x_640_ptq [6]	Coral	–	–
yolov5s [7]	Coral	139.87	2.89

Table 7. Continuation.

Raspberry Pi 5 Input size 640x640			
Basic models			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	TfLite	285.00	4.92
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	TfLite	79.03	3.10
efficientdet_lite3x_640_ptq [6]	TfLite	188.83	5.76
yolov5s [7]	TFLite	67.93	3.75
SW-optimized			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	ARM NN Delegate	355.26	12.90
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	ARM NN Delegate	193.99	6.34
efficientdet_lite3x_640_ptq [6]	ARM NN Delegate	304.65	13.69
yolov5s [7]	ARM NN Delegate	74.70	3.85
HW accelerated			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	Coral	359.94	5.90
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	Coral	130.07	2.90
efficientdet_lite3x_640_ptq [6]	Coral	-	-
yolov5s [7]	Coral	139.51	2.10
Nvidia Jetson Nano 2GB Input size 640x640			
Basic models			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	TfLite	2133.95	42.14
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	TfLite	238.39	4.66
efficientdet_lite3x_640_ptq [6]	TfLite	704.94	8.20
yolov5s [7]	TFLite	343.52	4.94
SW-optimized			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	ARM NN Delegate	2107.91	37.87
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	ARM NN Delegate	524.02	7.06
efficientdet_lite3x_640_ptq [6]	ARM NN Delegate	1044.74	19.04
yolov5s [7]	ARM NN Delegate	379.24	8.84
HW accelerated			
tf2_ssd_mobilenet_v1_fpn_640x640_coco17_ptq [6]	TensorRT	732.59	50.00
ssd_mobilenet_v2_fpnlite_640x640_coco17 [6]	TensorRT	141.88	16.59
efficientdet_lite3x_640_ptq [6]	TensorRT	-	-
yolov5s [7]	TensorRT	109.23	4.82

A graphic representation of the received results is presented in Fig. 1. As represented in Table 7 and Figures 1(a), (b), (c), and using an academic representation of real-time execution (100 ms at most or one fps), real-time execution is reached only for Raspberry Pi 5 target device, which has more CPU performance in comparison to other devices. Real-time execution is not attained for most models with the baseline list. Namely, such an approach is used for less powerful devices with ARM Cortex-A processors. So, base models of large input sizes are unsuitable for such devices.

Samples optimization approach for real-time object detection on single-board computers

Partial recognition is an approach that can be applied to solve problems with long execution times for devices with smaller computational resources. The main idea of such a method is to use several models with smaller input sizes to reduce the computational complexity of the overall task.

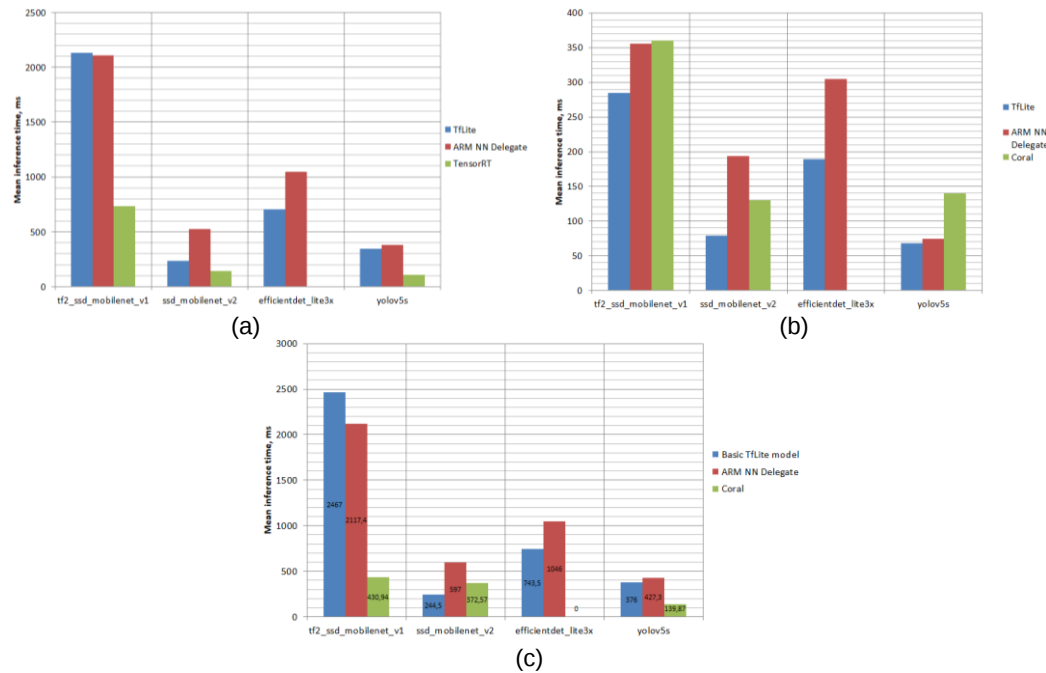


Fig. 1. The time comparison inference for models with input size 640x640 for target devices: Raspberry Pi 5 (a), Raspberry Pi 4 (b), Jetson Nano 2 GB (c).

Reference models with reduced (320x320 RGB images (Fig. 2) are used to confirm the real-time execution approach. The popular object detection model set is used for such an approach. All models were part of the standard TensorFlow Zoo repository and Google Coral model Zoo for external NPU-optimized models. TensorFlow standard model Zoo models, compiled with the TensorRT compiler, are used for the Jetson Nano target as an NPU-optimized case. Table 8 and Figure 3 (a), (b), and (c) present the inference results for 320x320 input size models.

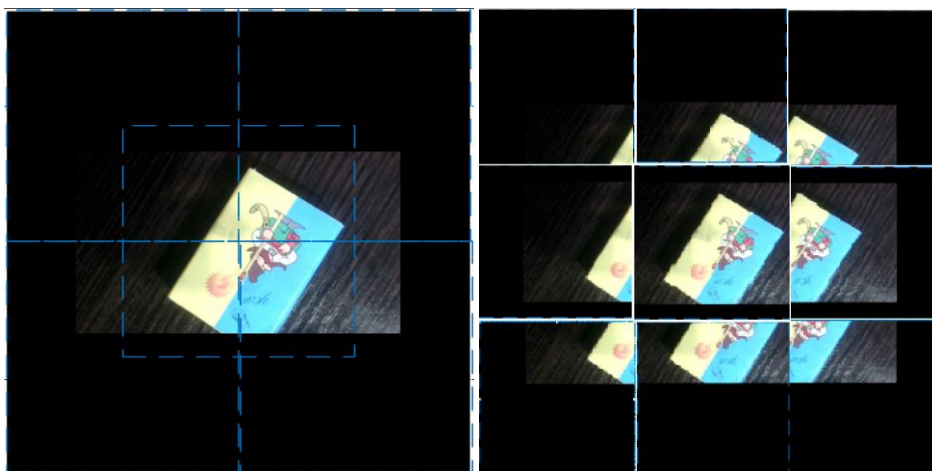


Fig. 2. Visual representation of image portioning.

Table 8. Models speed inference estimation with input size 320x320 RGB on target devices.

Model name	Model optimization	$\bar{x}$, ms	σ , ms
Raspberry Pi 4 Input size 320x320			
Basic models			
mobilenet_v1 [14]	TFLite	70.40	2.85
mobilenet_v2[14]	TfLite	40.69	0.79
efficientdet_lite0 [14]	TfLite	79.82	2.40
yolov5n [7]	TfLite	29.90	1.30
yolov5s [7]	TfLite	90.75	1.54
SW-optimized			
mobilenet_v1 [14]	ARM NN Delegate	68.88	7.19
mobilenet_v2 [14]	ARM NN Delegate	61.10	6.58
efficientdet_lite0 [14]	ARM NN Delegate	147.74	6.92
yolov5n[7]	ARM NN Delegate	47.42	2.77
yolov5s [7]	ARM NN Delegate	109.40	7.19
HW accelerated			
mobilenet_v1 [14]	Coral	10.37	1.88
mobilenet_v2[14]	Coral	15.56	1.94
efficientdet_lite0 [14]	Coral	89.68	4.96
yolov5n [7]	Coral	10.84	0.52
yolov5s [7]	Coral	18.86	1.86
Raspberry Pi 5 Input size 320x320			
Basic models			
mobilenet_v1 [14]	TFLite	10.50	0.31
mobilenet_v2[14]	TfLite	9.29	1.17
efficientdet_lite0 [14]	TfLite	22.76	2.15
yolov5n [7]	TfLite	6.30	0.56
yolov5s [7]	TfLite	14.71	0.83
SW-optimized			
mobilenet_v1 [14]	ARM NN Delegate	11.27	3.45
mobilenet_v2[14]	ARM NN Delegate	13.11	2.75
efficientdet_lite0 [14]	ARM NN Delegate	47.83	4.93
yolov5n [7]	ARM NN Delegate	12.81	1.68
yolov5s [7]	ARM NN Delegate	20.24	2.19
HW accelerated			
mobilenet_v1 [14]	Coral	9.94	2.19
mobilenet_v2[14]	Coral	13.19	2.05
efficientdet_lite0 [14]	Coral	56.81	2.24
yolov5n [7]	Coral	10.81	0.63
yolov5s [7]	Coral	18.89	2.19
Nvidia Jetson Nano 2GB Input size 640x640			
Basic models			
mobilenet_v1 [14]	TFLite	71.14	0.94
mobilenet_v2[14]	TfLite	39.60	0.69
efficientdet_lite0 [14]	TfLite	77.61	2.25
yolov5n [7]	TfLite	31.43	0.38
yolov5s [7]	TfLite	88.35	1.15
SW-optimized			
mobilenet_v1 [14]	ARM NN Delegate	98.43	15.63
mobilenet_v2[14]	ARM NN Delegate	63.00	6.62

Table 8. Continuation.

efficientdet_lite0 [14]	ARM NN Delegate	155.65	5.69
yolov5n [7]	ARM NN Delegate	48.27	3.23
yolov5s [7]	ARM NN Delegate	110.20	11.30
HW accelerated			
mobilenet_v1 [14]	TensorRT	61.24	3.92
mobilenet_v2[14]	TensorRT	41.37	3.92
efficientdet_lite0 [14]	TensorRT	–	–
yolov5n [7]	TensorRT	22.67	5.55
yolov5s [7]	TensorRT	32.59	4.68

A graphic representation of the received results is presented in Figure 3. Real-time execution is reaching for most models and target devices. It is represented in Table 8 and Figures 3(a), 3(b), and 3(c), using an academic representation of real-time execution (100 ms at most or one fps). Such an approach is suitable for less powerful devices with ARM Cortex-A processors.

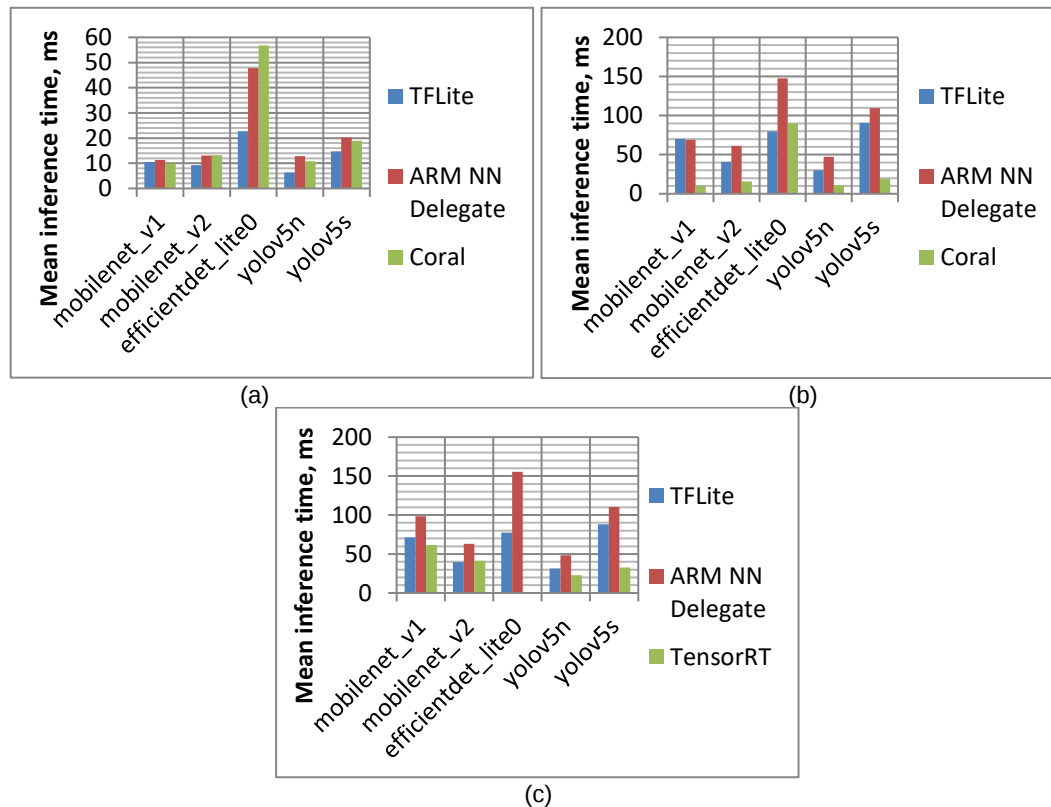


Fig. 3. Inference time comparison for models with input size 320x320 for target devices: Raspberry Pi 4 (a), Raspberry Pi 5 (b), Jetson Nano 2 GB (c).

CONCLUSION

As presented in tables 7 and 8 and diagrams 1 and 2, inferencing large accurate models with good accuracy is not reaching real-time execution on single-board computers based on Cortex A-series processors without optimizations. The popular object detection

models with image sizes 640x640 are mostly not suitable for such target devices but can provide more accurate results on most sizes of objects.

To resolve this problem, partial recognition with pre-processing can be applied, using models with less image size. Such type of models have significantly less computational complexity, but can provide good accuracy for large and medium object sizes. Moreover, preparation algorithms for this approach can be used in other parts of object detection and identification process. It can be helpful in complex systems with multi-stage pipelines for object detection and identification.

Such an approach was tested on partial object detection with an image size 320x320 for an original image size 640x640. Less complex models have provided significant inference time reduction. It can give real-time execution for most researched models and target devices, even by recognizing several interesting zones.

AUTHOR CONTRIBUTIONS

Conceptualization, [D.M.]; methodology, [D.M.]; validation, [B.B.]; formal analysis, [D.M.]; investigation, [D.M.]; resources, [D.M.]; data curation, [D.M.]; writing – original draft preparation, [D.M.]; writing – review and editing, [B.B.]; visualization, [D.M.] supervision, [B.B.]; project administration, [D.M.]; funding acquisition, [D.M.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Ltd, R. P. (Trading). (2023). Buy a Raspberry Pi 4 Model B. Raspberry Pi. <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/>
- [2] Raspberry Pi Ltd. (n.d.). Buy a Raspberry Pi 5. Raspberry Pi. <https://www.raspberrypi.com/products/raspberry-pi-5/>
- [3] NVIDIA Jetson Nano. (2024). NVIDIA. <https://www.nvidia.com/ru-ru/autonomous-machines/embedded-systems/jetson-nano/>
- [4] Dev Board. (n.d.). Coral. <https://coral.ai/products/dev-board/>
- [5] MLPerf Inference - MLCommons. (2024, November 18). MLCommons. <https://mlcommons.org/working-groups/benchmarks/inference/>
- [6] Tensorflow. (2019, July 15). Tensorflow/models. GitHub. https://github.com/tensorflow/models/tree/master/research/object_detection
- [7] Jocher, G. (2020, August 21). ultralytics/yolov5. GitHub. <https://github.com/ultralytics/yolov5>
- [8] Ameen, S., Siriwardana, K., & Theodoridis, T. (2023). Optimizing Deep Learning Models For Raspberry Pi. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2304.13039>
- [9] Assunção, E., Pedro Dinis Gaspar, Khadijeh Alibabaei, Maria Paula Simões, Proença, H., Vasco, & Caldeira, P. (2022). Real-Time Image Detection for Edge Devices: A Peach Fruit Detection Application. Future Internet, 14(11), 323–323. <https://doi.org/10.3390/fi14110323>
- [10] Gabriele Proietti Mattia, & Beraldi, R. (2021). A study on real-time image processing applications with edge computing support for mobile devices. IRIS Research Product Catalog (Sapienza University of Rome). <https://doi.org/10.1109/ds-rt52167.2021.9576139>
- [11] Eduardo Timóteo Assunção, Pedro Dinis Gaspar, Ricardo Alves Mesquita, Maria João Simões, Khadijeh Alibabaei, André Veiros, & Proença, H. (2022). Real-Time Weed Control Application Using a Jetson Nano Edge Device and a Spray Mechanism. 14(17), 4217–4217. <https://doi.org/10.3390/rs14174217>
- [12] Post-training quantization | TensorFlow Lite. (n.d.). TensorFlow. <https://www.tensorflow.org/lite/performance>

- [13] COCO - Common Objects in Context. (n.d.). Cocodataset.org.
<https://cocodataset.org/>
- [14] Models - Object Detection | Coral. (2020). Coral; Coral.
<https://coral.ai/models/object-detection/>

ОПТИМІЗАЦІЇ МОДЕЛЕЙ ДЕТЕКТУВАННЯ ОБ'ЄКТІВ ДЛЯ ПРИСКОРЕННЯ ЇХ ВИКОНАННЯ НА ОДНОПЛАТНИХ ПЛАТФОРМАХ ЗАГАЛЬНОГО ПРИЗНАЧЕННЯ ТА ПЛАТФОРМАХ З АПАРАТНИМ ПРИСКОРЕННЯМ

Дмитро Миронюк, Богдан Благітко

*Львівський національний університет імені Івана Франка
 вул. Ген. Тарнавського, 107, Львів, Україна*

АНОТАЦІЯ

Вступ. Надано опис та підготовку сучасних підходів глибинного навчання до моделей виявлення об'єктів. В якості основи використовуються фреймворки глибинного навчання для навчання моделей і висновків, такі як TensorFlow і TensorFlow Lite. Проаналізовано концепції оптимізації моделі глибинного навчання. Квантовані моделі int8 використовуються як базові для оцінки ефективності оптимізації. Підхід делегування включає оптимізовані програмні або апаратні варіанти нейронних операцій. Він підготований для прискорення процесу виявлення об'єктів на цільових пристроях.

Матеріали та методи. Пристрій зі зниженими ресурсами продуктивності або мікроконтролер без блоків з плаваючою комою використовує модель базової оптимізації з вагами int8. Фреймворк TensorFlow Lite має різні типи квантування, викладені в детальному поясненні. Підготовані бенчмарки для сучасних одноплатних пристроїв, а також здійснені кореляції між використанням різних підходів до оптимізації, типів одноплатних платформ і аналіз швидкості виявлення об'єктів. Усі протестовані моделі попередньо навчені з використанням набору даних MS COCO (80 класів). Усі моделі були підготовлені для експерименту з 8-бітним повним цілочисельним квантуванням і генерацією вихідної моделі TFLite з використанням зображень TensorFlow Object Detection API Docker і Python 3.11.

Результати. Зразки даних тестування отримано з архіву набору даних перевірки MS COCO. Розмір вхідного зображення 640x640 RGB. Порівняння часу розпізнавання зображення з 640x640 RGB проводилося на Raspberry Pi 5, Raspberry Pi 4 і Jetson Nano 2GB. Лише цільовий пристрій Raspberry Pi 5 досяг виконання в реальному часі (не більше 100 мс або один кадр/с), оскільки він має більшу продуктивність ЦП, ніж інші пристрої. Підтвердження виконання в режимі реального часу було досягнуто за допомогою еталонних моделей зі зменшеними розмірами зображення (320x320 RGB).

Висновки. Стандартні моделі TensorFlow. Моделі Zoo, скомпільовані за допомогою компілятора TensorRT, використовувалися для Jetson Nano як варіант, оптимізований для NPU. Виконання в режимі реального часу (не більше 100 мс або один кадр в секунду) досягається на більшості моделей і цільових пристроїв. Такий підхід доцільно використовувати на менш потужних пристроях з процесорами ARM Cortex-A.

Ключові слова: одноплатні комп'ютери, моделювання, контрольна таблиця, нейронні мережі, детектування об'єктів, оптимізація.

УДК: 004.8; 338.2

ПРОБЛЕМИ ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРОГНОЗУВАННЯ ЦІНИ ВІРТУАЛЬНИХ АКТИВІВ

Андрій Цемко¹, Максим Мацьків²

Львівський національний університет імені Івана Франка

¹ Кафедра радіофізики та комп'ютерних технологій,
вул. Ген. Тарнавського, 107, 79017 Львів, Україна

² Кафедра системного проектування,
вул. Драгоманова, 50, 79005 Львів, Україна

Цемко, А.М., Мацьків, М.А. (2025). Проблеми застосування нейронних мереж для прогнозування ціни віртуальних активів. *Електроніка та інформаційні технології*, 29, 69-78.
<https://doi.org/10.30970/eli.29.7>

АНОТАЦІЯ

Вступ. Прогнозування цін віртуальних активів є важливим завданням через їх високу мінливість. Нейронні мережі широко використовуються для таких задач, але часто стикаються з проблемою наївних передбачень, коли наступне значення надто схоже на попереднє, що знижує ефективність прогнозування.

Матеріали та методи. Дослідження проводилося на основі даних динаміки цін Bitcoin (біржа Coinbase) за період 4 травня 2021 – 9 квітня 2024. Розглянуто два методи нормалізації даних цін: лінійну (min-max) та нормалізацію співвідношення (ratio), яка забезпечує стаціонарність даних. Нейронна мережа із двонаправленими LSTM-шарами навчалася на 10 попередніх значеннях цін для прогнозування одного наступного. Використано бібліотеку TensorFlow, зокрема інтерфейс Keras, а процес навчання складався з: 150 навчальних епох, тренувальних вибірок по 32 навчальні пари, оптимізатор Adam і функція втрат MSE.

Результати й обговорення. Лінійна нормалізація показала найгірші результати, оскільки модель втрачає здатність до прогнозування, якщо майбутні значення виходять за межі навчальної вибірки. Натомість нормалізація відношення продемонструвала значно кращі результати, дозволяючи моделі враховувати динаміку змін навіть за межами мінімальних і максимальних значень навчальних даних. Використання додаткового множника для нормалізованих даних ($k = 400$) покращило результати для тренувальних даних, хоча на тестовому наборі покращення було незначним. Додавання додаткових характеристик цін (під час відкриття, закриття торгів та мінімальна ціна протягом цього періоду) не змогло усунути проблему зсуву прогнозів. Аналіз результатів показав, що модель часто базується на повторенні попередніх значень, що свідчить про її обмеженість у захопленні складних взаємозв'язків у даних.

Висновки. Дослідження підтверджує, що нейронні мережі мають обмеження у задачах прогнозування цін віртуальних активів. Проблема наївного передбачення є ключовою перешкодою, яка обмежує ефективність моделей. Використання нормалізації відношення із множником k покращує точність, але цього недостатньо для вирішення основної проблеми. Необхідні нові підходи чи вдосконалення існуючих методів, що дозволять моделі краще розуміти складні взаємозв'язки у часових рядах і надійно прогнозувати майбутні зміни.

Ключові слова: рекурентні нейромережі, прогнозування, віртуальні активи.



© 2025 Andrii Tsemko & Maxym Matskiv. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ВСТУП

Метою роботи є висвітлення проблеми використання нейронних мереж для прогнозування майбутніх значень ціни віртуальних активів базуючись на динаміці їхньої зміни за попередні інтервали часу. Під час дослідження застосування нейронних мереж було виявлено фундаментальну проблему, яка потребує глибшого дослідження і висвітлення в задачах прогнозування часових рядів.

Опираючись на результати минулої статті [1], вирішено провести поглиблене дослідження методів нормалізації даних динаміки цін, які в подальшому будуть використані для навчання нейронних мереж. Проте під час цього дослідження виявлено проблему, коли нейронна мережа передбачає наступне значення, як дуже схоже на попереднє, що давалося на вхід (так зване передбачення “як вчора”) [2].

Провівши глибший аналіз літератури, ми побачили, що невеликий розмір нейронної мережі не є основною проблемою. Дана проблема передбачення часто виникає у схожих підходах з нейронними мережами для задач прогнозування. Опираючись на результати дослідження авторів статті “Are Transformers Effective for Time Series Forecasting?”, складніші структури нейронних мереж такі як трансформери, показують точність прогнозуванням на тому самому рівні як і простіші моделі [3]. Тому аспект складності нейронної мережі для задачі прогнозуваннями нами досліджувався лише поверхнево.

Для ефективного процесу навчання нейронних мереж варто використовувати нормалізовані дані. Одним з найпоширенішим методом нормалізації є метод лінійної нормалізації (min-max) [4]. Зазвичай, в цій нормалізації значення зводяться до проміжку від 0 до 1, де 0 – це мінімальне значення (min), а 1 – це максимальне (max). Формула лінійної нормалізації даних:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (1)$$

де X – оригінальні дані;

X_{min} та X_{max} – мінімальне та максимальне значення оригінальних даних відповідно;

X_{norm} – нормалізовані дані.

Незважаючи на результати навчання нейронної мережі, використання лінійної нормалізації має недолік, який потенційно може зменшити ефективність застосування нейронної мережі для прогнозування подальших змін ціни. У випадку, якщо майбутні значення цін вийдуть за межі мінімального і максимального значень даних, використаних у навчанні, нейронна мережа не зможе спрогнозувати подальшу зміну цін, оскільки вона не бачила таких значень. Цю проблему було виявлено за минулого дослідження, що відображено на Рис. 1.

Для вирішення цієї проблеми запропоновано використати інший підхід нормалізації даних – нормалізації відношення (ratio). Її суть полягає в зображенні даних як відношення наступного значення $n+1$ до поточного n (або поточного n до попереднього $n-1$). Такий підхід нормалізації також зустрічається в статтях під назвою відсоткова нормалізація. Дані після застосування такої нормалізації знаходяться у проміжку між 0,85 і 1,15, де значення менше 1 відповідає зниженню ціни, а значення більше 1 – зростанню. Таке представлення даних є стаціонарним з середнім значенням в околі 1, що дозволяє використовувати навчену нейронну мережу для прогнозування майбутніх даних, навіть при виході даних за оригінальні мінімальні і максимальні значення, які були використані під час навчання нейронної мережі (Рис. 2).

Отримані дані можна охарактеризувати як стаціонарний процес, оскільки середнє значення не змінюється з часом. Для підвищення ефективності навчання



Рис. 1. Порівняння прогнозованих значень нейронної мережі з оригінальними. Синя лінія – навчальні дані; помаранчева – тестувальні; зелена – прогнозовані значення.

Fig. 1. Comparison of the predicted values of the neural network with the original ones. Blue line – training data; orange – testing data; green – predicted data.

нейронної мережі, ми використовуємо додатковий множник для нормалізованих даних, суть якого полягає у тому, щоб збільшити середнє квадратичне відхилення. Формула нормалізації відношення має вигляд:

$$X_{norm} = \left(\frac{X_n}{X_{n-1}} - 1 \right) \cdot k, \quad (2)$$

де k – множник нормалізації.

Метою даного дослідження є аналіз ефективності альтернативних підходів нормалізації даних у задачах прогнозування часових рядів на основі нейронних мереж для вирішення виявлених проблем точності та адаптивності моделей.

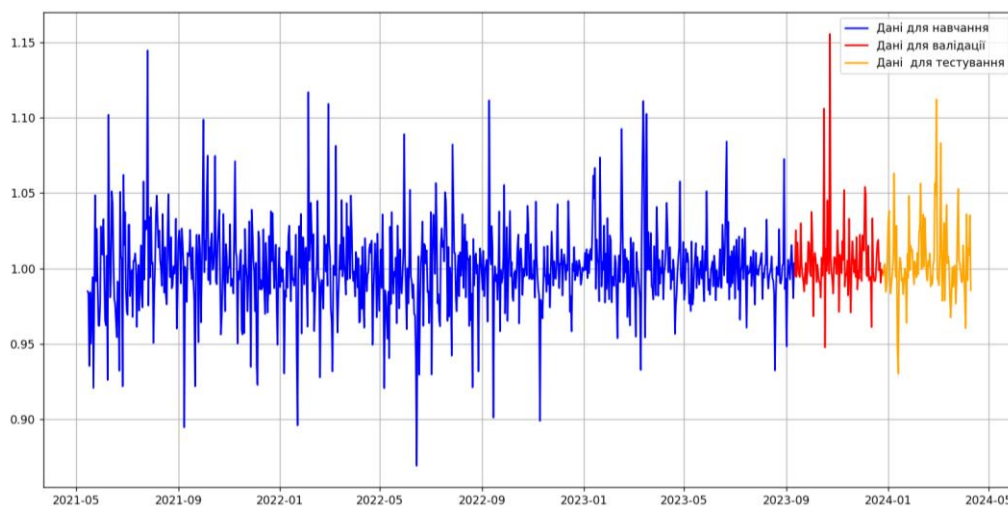


Рис. 2. Дані після нормалізації відношення. Синя лінія – навчальні дані; помаранчева – тестувальні; червона – дані перевірки.

Fig. 2. Ratio normalized data. Blue line – training data; orange – testing data; red – validation data.

МАТЕРІАЛИ ТА МЕТОДИ

Всі описані результати дослідження в цій статті були виконані з використанням даних динаміки цін криптовалюти Bitcoin відносно долару США. Ці дані було взято з біржі Coinbase [5]. Для дослідження було використано дані в часовому проміжку від 4 травня 2021 року до 9 квітня 2024 року, а саме щоденне співвідношення вартості обраних активів (Рис. 3).

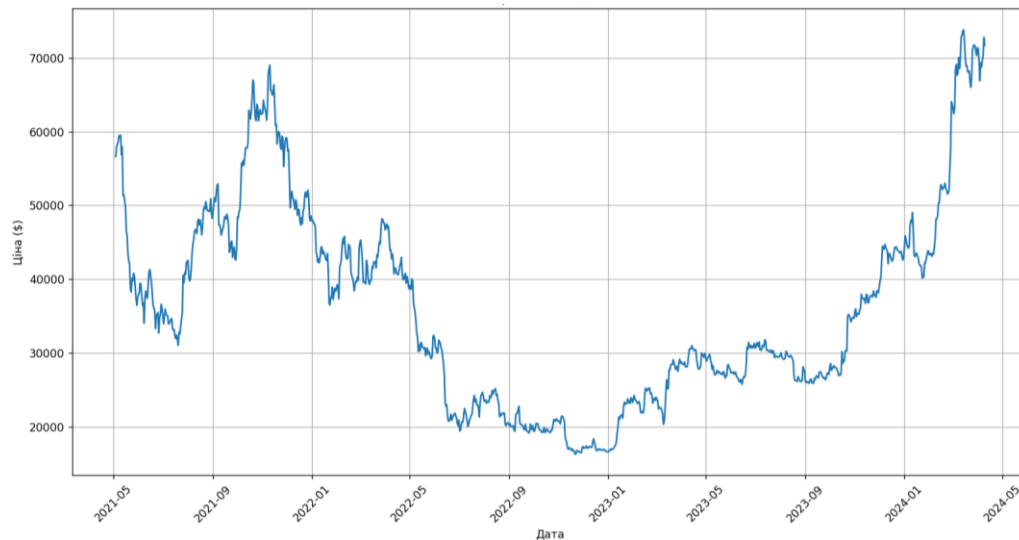


Рис. 3. Максимальна вартість Bitcoin'у по днях з біржі Coinbase
Fig. 3. Maximum daily Bitcoin price from the Coinbase exchange

Для створення та навчання нейронної мережі, було використано бібліотеку TensorFlow [6] мови програмування Python від компанії Google, з відкритим кодом. Бібліотека TensorFlow містить в собі інтерфейс Keras, який дозволяє створювати структури нейронних мереж пошарово, налаштовуючи відповідні параметри кожного з шарів. Так, для створення рекурентних нейронних мереж інтерфейс Keras надає можливість створення таких шарів: RNN (Recurrent Neural Networks, Рекурентна нейронна мережа), LSTM (Long Short-Term Memory, Довга короткочасна пам'ять) та GRU (Gated Recurrent Unit, Вентильний рекурентний вузол) [7, 8].

У моделі нейронної мережі використовується архітектура з двонаправленими LSTM-шарами, що дозволяють враховувати контекст як з попередніх, так і з наступних кроків послідовності. Перший шар містить 50 нейронів передає всю послідовність на наступний шар з функцією активації ReLU (Rectified Linear Unit, Випрямлений лінійний вузол або Випрямляч). Далі йде шар вимкнення 20% нейронів для запобігання перенавантаженню. Другий двонаправлений шар LSTM містить 54 нейрони та вже не повертає послідовність. Потім додаються два щільні (Dense) шари: перший з 50 нейронами і функцією активації ReLU, другий з 25 нейронами без функції активації. Після кожного з них також відбувається вимкнення 20% нейронів для регуляризації. Після останнього вихідного шару нейронна мережа передає згенероване наступне значення (Рис. 4).

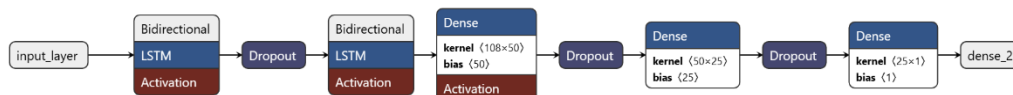


Рис. 4. Структура нейронної мережі.
Fig. 4. Neural network structure.

Для оцінки точності прогнозування моделі використовувався метод середньоквадратичної похибки (Mean Squared Error, MSE). Дані попередньо нормалізувалися за методами, детально описаними у попередньому розділі.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

Для дослідження було вирішено навчити нейронну мережу, яка буде генерувати прогноз наступного значення ціни на основі динаміки цін за останні 10 днів. Навчання нейронної мережі проводилось протягом 150 епох, з використання вибірок по 32 набори тренувальних пар. Використано дані за 1072 дні, які було розділено у співвідношенні 80% до 20% на дані для навчання та верифікації. В якості оптимізатора ваг нейронної мережі використано оптимізатор Adam і функцію обрахунку втрат – середньоквадратичну похибку (Mean Squared Error, MSE). Під час процесу навчання для функції втрат використовується значення похибки нормалізованих даних. Результати навчання (наведені у таблиці 1) обраховуються на основі денормалізованих даних.

Таблиця 1. Результати навчання нейронної мережі прогнозування одного значення на основі 10-ти вхідних параметрів лінійно нормалізованих даних.
Table 1. Neural network training results predicting a single value based on 10 input parameters linearly normalized data.

	Навчальні дані	Тестувальні дані
Середньоквадратична похибка (\$)	2,29 млн	7,35 млн

Формула денормалізації лінійно нормалізованих даних має вигляд:

$$X_{denorm} = X_{norm} * (X_{max} - X_{min}) + X_{min}, \quad (3)$$

а формула для денормалізації даних для нормалізації відношення:

$$X_{denorm} = \left(\frac{X_{norm}}{k} + 1 \right) * X_{n-1}. \quad (4)$$

Нами було досліджено вплив значення множника k з формули (2) на результати навчання нейронної мережі. Було досліджено ефективність навчання нейронної мережі при значення множника 1, підбраного значення 36, що відповідає представленню даних стаціонарного процесу характеристикою якого є значення середньоквадратичного відхилення (СКВ) рівне 1 та значення множника 400, що показує кращі результати. Результати середньоквадратичних похибок (MSE) після денормалізації даних занесені до таблиці 2.

Аналізуючи результати таблиці 2, ми бачимо, лінійна нормалізація показує найгірший результат, а нормалізація відношення демонструє значення похибки майже в 4 рази менше для тестової вибірки. Множник 400 для нормалізації відношення був найоптимальнішим для навчальної вибірки.

Аналізуючи графіки на рис. 5 та 6 можна побачити, що всі способи нормалізації відтворюють оригінальні дані зі зсувом вправо, а лінійна нормалізація також демонструє певний зсув вниз. Згідно значень похибки, використання представлення даних з нормалізацією відношення показує в два рази менше значення похибки, а використання різних коефіцієнтів лише покращує значення для тренувальних даних, в той час як значення на тестовому наборі майже не змінюються. Це свідчить про те, що нейронна мережа вчиться "як вчора".

Таблиця 2. Результати середньоквадратичних похибок для лінійної нормалізації та нормалізації відношення.

Table 2. Mean squared error results for linear and ration normalization.

Тип нормалізації	Тренувальні дані	Тестові дані
Лінійна нормалізація	2,3 млн	7,35 млн
Нормалізація відношення з множником 1	1,1 млн	2,09 млн
Нормалізація відношення з множником 36 (стандартне відхилення=1)	0,3 млн	2,88 млн
Нормалізація відношення з множником 400	0,03 млн	2,81 млн

Оскільки нормалізація відношення з множником 400 показала найкращий результат, було вирішено її й залишити. Минулі дослідження використовували тільки максимальне значення ціни протягом часового відтинку. Прийнято рішення створити



Рис. 5. Результати прогнозування нейронних мереж з нормалізаціями відношення: а) весь діапазон даних; б) тренувальна вибірка; в) тестова вибірка.

Fig. 5. Prediction results of neural networks with ratio normalizations: (a) the entire data range; (b) training set; (c) testing set.

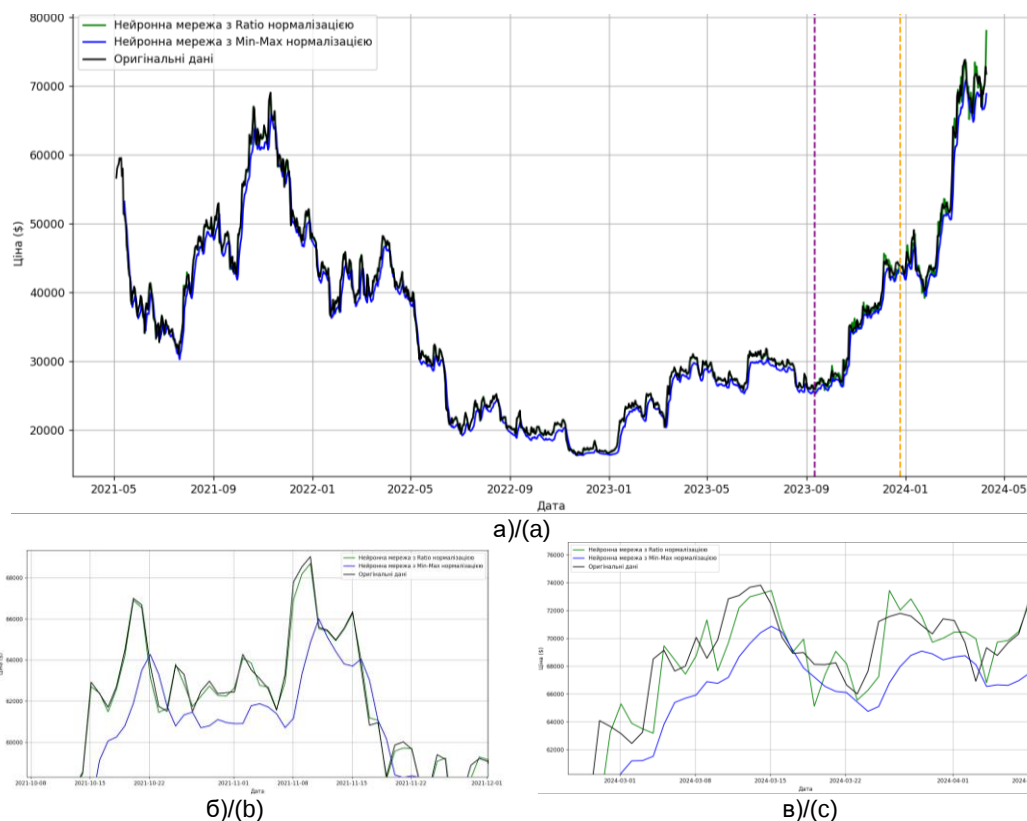


Рис. 6. Результати прогнозування нейронних мереж з лінійною та нормалізацією відношення: а) весь діапазон даних; б) тренувальна вибірка; в) тестова вибірка.

Fig. 6. Prediction results of neural networks with min-max and ratio normalizations: (a) the entire data range; (b) training set; (c) testing set.

нейронну мережу, яка не буде обмежуватись тільки максимальними значеннями ціни, а також використовувати інші характеристики, такі як значення ціни відкриття, закриття та мінімуму протягом часового відліку. Для цього використовуються ті самі дані з біржі Coinbase (за період від 4 травня 2021 року до 9 квітня 2024 року). Нами було створено нейронну мережу, яка обробляє масив значення за 10 днів, який містить 4 колонки значень: максимальної і мінімальної цін, та цін відкриття (початку торгів) і закриття (закінчення торгів), і прогнозує максимальну ціну активу на наступний день. Рис. 7 відображає результати навчання такої нейронної мережі. Дані на графіку поділені на тренувальну, контрольну та тестову вибірки. Дані від початку і до фіолетової прямої відповідають тренувальній вибірці, між фіолетовою та жовтою – контрольні дані, які використовуються для контролю процесу навчання для уникнення перенавчання нейронної мережі і тестові дані.

З графіка видно, що збільшення кількості характеристик не допомогло вирішити проблему з прогнозуванням, оскільки результати прогнозування візуально ніби відстають на 1 часовий відлік, що можна охарактеризувати як тип прогнозування «завтра буде як вчора». Для кращого розуміння проблеми, в таблиці 3 порівнюються значення середньоквадратичних похибок для прогнозованих даних як вони є (з відставанням), та прогнозованих даних, зміщених на один часовий відлік вліво.

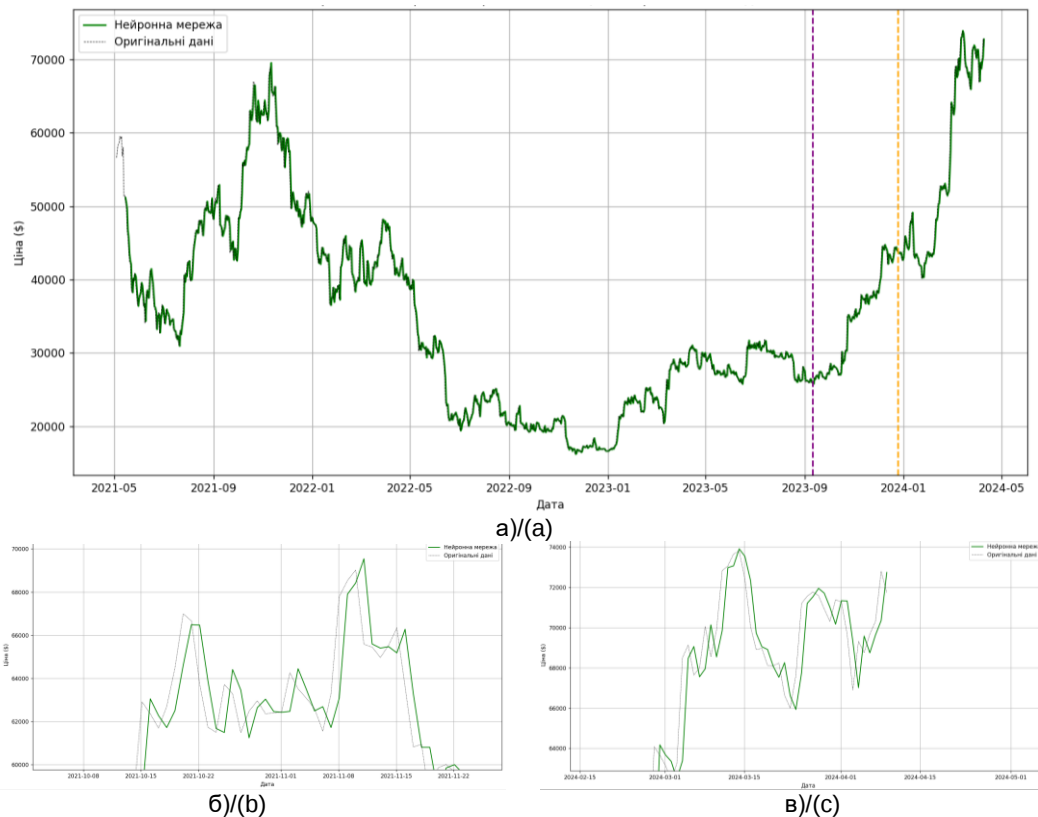


Рис. 7. Результати прогнозування нейронної мережі з нормалізацією відношення, коли на вхід беруться 4 колонки: а) весь діапазон даних; б) тренувальна вибірка; в) тестова вибірка.
Fig. 7. Prediction results of the neural network with ratio normalization when using 4 columns as input: (a) the entire data range; (b) training set; (c) testing set.

Таблиця 3. Порівняння значень середньоквадратичних похибок оригінальних прогнозованих даних та прогнозованих даних після їх зміщення на один часовий відлік вліво.
Table 3. Comparison of mean squared errors of the original forecasted data and the forecasted data after shifting one time step to the left.

Тип даних	Тренувальні дані	Тестові дані
Оригінальні прогнозовані	1,06 млн \$	2,02 млн \$
Прогнозовані дані зміщення на один часовий відлік вліво	0,01 млн \$	0,01 млн \$

ВИСНОВКИ

Таким чином, аналізуючи результати досліджень можна вважати, що використання нейронних мереж для прогнозування ціни віртуальних активів не є ефективним. Така мережа, хоч у певних випадках і показує хороші результати для тренувальних даних, але все ж для тестових даних обирає найкраще рішення наступного елемента як оригінальне попереднє. Нейромережа не справляється із задачею прогнозування ціни, оскільки її передбачення базуються на повторенні

попередніх значень замість справжнього прогнозування. Це свідчить про обмеження моделі в її здатності захоплювати складні взаємозв'язки у даних, що призводить до необхідності пошуку інших підходів або вдосконалення існуючих методів для покращення прогнозування.

ВНЕСОК АВТОРІВ

Концептуалізація, [А.Ц., М.М.]; методологія, [А.Ц., М.М.]; валідація, [М.М.]; формальний аналіз, [А.Ц., М.М.]; дослідження, [А.Ц., М.М.]; написання – підготовка оригінального тексту, [М.М.]; написання – рецензування та редагування, [А.Ц.]; візуалізація, [М.М.] нагляд, [А.Ц.]; адміністративне управління проектом, [А.Ц.].

Усі автори ознайомилися з остаточною версією рукопису та погодилися з його публікацією.

ДЖЕРЕЛА

- [1] Tsemko, A., & Liubun, Z. (2022). USING A NEURAL NETWORK FOR PRICE PREDICTION OF VIRTUAL ASSETS. Electronics and Information Technologies, 20. <https://doi.org/10.30970/eli.20.5>
- [2] Erharter, G. H., & Marcher, T. (2021). On the pointlessness of machine learning based time delayed prediction of TBM operational data. Automation in Construction, 121, 103443. <https://doi.org/10.1016/j.autcon.2020.103443>
- [3] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121–11128. <https://doi.org/10.1609/aaai.v37i9.26317>
- [4] Kim, Y.-S., Kim, M. K., Fu, N., Liu, J., Wang, J., & Srebric, J. (2024). Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using different Artificial Neural Network Models. Sustainable Cities and Society, 105570. <https://doi.org/10.1016/j.scs.2024.105570>
- [5] Coinbase. (n.d.). Печуц: Coinbase API documentation. Retrieved from <https://docs.cdp.coinbase.com>
- [6] TensorFlow. (2023). Печуц: TensorFlow guide. Retrieved from <https://www.tensorflow.org/guide>
- [7] TensorFlow. (2024). Печуц: Time series tutorial. Retrieved from https://www.tensorflow.org/tutorials/structured_data/time_series
- [8] Brownlee, J. (2022, August 7). Time series prediction with LSTM recurrent neural networks in Python with Keras. Retrieved from <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>

PROBLEMS OF USING NEURAL NETWORKS TO PREDICT THE PRICE OF VIRTUAL ASSETS

Andrii Tsemko¹  & Maxym Matskiv² 

Ivan Franko National University of L'viv

¹ *Department of Radiophysics and Computer Technologies,*

¹107 Gen. Tarnavskoho St., 79017 Lviv, Ukraine

² *Department of System Design,*

50 Drahomanova St., 79005 Lviv, Ukraine

ABSTRACT

Background. Predicting the prices of virtual assets is an important task due to their high volatility. Neural networks are widely used for such tasks, but often face the problem of naive predictions, when the next value is too similar to the previous one, which reduces the forecasting efficiency.

Materials and Methods. The study was conducted on the basis of Bitcoin price dynamics data (Coinbase exchange) for the period of May 4, 2021 - April 9, 2024. Two methods of data normalization were considered: linear (Min-Max) and ratio normalization, which ensures data stationarity. A neural network with bidirectional LSTM layers was trained on 10 previous values to predict one subsequent value. The TensorFlow library, in particular the Keras API, was used for training process with customizable parameters: 150 epochs, 32 batch size, Adam optimizer, and MSE loss function.

Results and Discussion. Linear normalization showed the worst results, as the model loses its ability to predict if future values fall outside the training set. On the contrary, ratio normalization showed much better results, allowing the model to take into account the dynamics of changes even beyond the minimum and maximum values of the training data. Using an additional multiplier for the normalized data ($k = 400$) improved the results for the training data, although the improvement was not significant on the test set. Adding additional price characteristics (at the opening, closing, and minimum price during this period) failed to eliminate the problem of forecast bias. The analysis of the results showed that the model is often based on the repetition of previous values, which indicates its limitations in capturing complex relationships in the data.

Conclusion. The study confirms that neural networks have limitations in the task of predicting virtual asset prices. The problem of naive prediction is a key obstacle that limits the effectiveness of models. The use of ratio normalization with a k -factor improves accuracy, but it is not enough to solve the main problem. New approaches or improvements to existing methods are needed to allow the model to better understand complex relationships in time series and reliably predict future changes.

Keywords: recurrent neural networks, forecasting, virtual assets.

UDC 004.78

ENERGY CONSERVATION AS ONE OF THE COMPONENTS OF THE MANAGEMENT SYSTEM FOR THE SMART SUSTAINABLE WORKSPACES

Maksym Yakubovych*, Vasyl Lyashkevych, Roman Shuvar

Department of System Design,
Ivan Franko National University of Lviv
50 Drahomanova St., UA-79005 Lviv, Ukraine

Yakubovych, M.Y., Lyashkevych, V.L., & Shuvar, R.Y. (2025). Energy Conservation as One of the Components of the Management System for the Smart Sustainable Workspaces. *Electronics and Information Technologies*, 29, 79–94. <https://doi.org/10.30970/eli.29.8>

ABSTRACT

Background. "Smart technologies" have grown rapidly thanks to advanced achievements in artificial intelligence (AI) and IoT. Big companies are interested in the concept of "smart rooms" or "smart workplaces" because these technologies allow them to reduce the costs of maintaining workplaces and focus on sustainable production without redundant waste.

Energy conservation is a crucially important part of a smart sustainable workplace management system therefore it has been analyzed in this paper. Considering the features of smart workplaces we highlighted that this approach offers a flexible and convenient way to make the workplace more comfortable and productive with a more accurate device management strategy.

Materials and Methods. Aiming our goals we have investigated the tasks which were related to models and methodologies for energy consumption management based on accessible technical resources and standards. Thus, we have used a systematic approach to selecting material, methods of inductive and logical analysis, observation and so on. The appropriate sensors and devices for energy management in smart workplaces were considered as well.

Results and Discussion. We paid the most attention to the devices available to us, such as Google Home, Mi Home, and Domoticz and their predefined functional characteristics. Amidst models, the most widespread ones for energy management were reviewed. Specifically noting that energy management in smart workplaces based on fuzzy logic does not require a complex mathematical model for system management and can rely directly on the experience of qualitative users.

Conclusion. Taking into account other components of a smart sustainable working space management system, we can conclude that modern information technologies and data analytics are becoming powerful tools for optimizing energy consumption and enhancing comfort and productivity in working spaces.

Keywords: smart sustainable workplace, smart sustainable workplace management system, energy consumption, energy conservation, adaptive-intelligent management.

INTRODUCTION

The implementation of advanced information technologies and data analytics is becoming a key factor in developing effective energy conservation methods. Data collection and analysis systems, such as "smart homes" or "smart grids", allow for



© 2025 Maksym Yakubovych et al. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and Information Technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

monitoring and optimizing energy consumption in real time. Buildings where software takes care of security, energy conservation, and resident comfort are called “smart” [1]. Thanks to IoT and data transmission technologies, household appliances are integrated into a unified system, which enables [2]:

- autonomous lighting control, turning off lights in empty rooms;
- monitoring the condition of water pipes and air quality, warning about possible leaks and contamination;
- receiving notifications about the risk of breakdowns.

Currently, energy management systems face rather stringent requirements, as excessive energy consumption leads to excessive consumption of planetary resources. [15] For these considerations, a number of requirements and recommendations are being developed that must be taken into account when planning workspaces. [4] Therefore, in modern conditions, we are talking about smart sustainable workspaces. [2]

The implementation of such systems is based on the application of modern technologies that:

- increase employee productivity;
- ensure efficient and economical use of workspaces [17].

Automatic regulation of lighting, heating, ventilation, and air conditioning based on data analysis of energy load and consumption allows: reducing unnecessary costs and providing optimal working conditions with minimal energy consumption while minimizing the use of natural resources.

Initial understanding of a sustainable workspace [5] is focused primarily on the responsible use of resources. A sustainable workspace was defined as an environment where products and services are created in a manner that does not deplete irreplaceable resources without compromising the quality of the output. This definition is intentionally broad, applicable to diverse settings ranging from offices to production lines, emphasizing that the core principle of sustainability lies in the nature of the work itself and the materials involved. The central idea is to minimize reliance on resources that cannot be replenished, thereby ensuring the long-term viability of operations and reducing the ecological footprint.

Expanding on this foundational view, the concept of a sustainable workplace culture introduces a more comprehensive perspective. This approach emphasizes environmental, social, and economic mindfulness across all facets of business operations.

The key characteristic of a sustainable workspace is a strong emphasis on energy efficiency through the adoption of renewable energy, the use of efficient technologies, and the optimization of building design and operational practices. Waste reduction was another critical element, moving beyond basic recycling to embrace circular economy principles like reuse, upcycling, and the minimization of single-use items. The selection of sustainable materials, characterized by recycled content, renewability, low environmental impact, and non-toxic properties, was also a defining feature. Furthermore, the concept of a sustainable workspace increasingly incorporated a focus on fostering employee well-being through factors like natural light, biophilic design, healthy materials, and supportive work environments. Finally, the social dimension extended to encompass the broader community through ethical practices, local engagement, and a commitment to corporate social responsibility.

Smart work environments: current state and terminology. Considering the current state and terminology in the field of smart spaces, we can establish that such work environments utilize various sensors, detectors, analytical tools, and the Internet of Things (IoT) for automation and optimization of different aspects of the work environment. These include:

- Lighting and energy conservation: in smart work environments, lighting systems are automatically adjusted depending on employee presence, natural light levels, and time of day. Additionally, systems for automatic shutdown of electrical appliances and equipment at night help reduce electricity consumption.

- Climate management: automated control systems for heating, ventilation, and air conditioning can respond to temperature changes and human presence in space, ensuring comfortable working conditions.
- Security systems: in smart work environments, video surveillance systems can use image analytics to detect suspicious activities or malfunctions [13].
- Workspace and zone reservation: Booking systems help employees find available workspaces, conference rooms, and other resources, as well as plan their workday.
- Cost management: analytical tools help track the use of resources such as electricity and water, and identify opportunities for savings.
- Health and comfort monitoring: sensors can measure air quality, humidity, and other parameters that affect employee comfort and health.
- Automatic equipment control: automation systems can provide automatic startup, shutdown, and regulation of equipment such as computers, printers, coffee makers, etc.

In addition to the above, the concept of a smart work environment can be interpreted as an intelligently constructed workspace. This means that when designing the building, it is necessary to consider the possibility of integrating all services with each other using wireless technologies. Such a high-tech system can combine all concepts into a single structure that can be conveniently managed even from a mobile phone.

Thus, a smart work environment refers to a high-technology workspace where integrated sensors, automation systems, artificial intelligence, and connectivity help provide optimal working conditions for employees, increase productivity, and reduce costs. To ensure the effective operation of a smart work environment, it is necessary to establish requirements for aspects such as lighting, ventilation, communication infrastructure, security, and privacy.

Researchers indicate that requirements for smart work environments should be flexible and adapted to the changing needs of employees.

A core tenet of the smart sustainable workspace is its commitment to sustainability. Scientific literature emphasizes resource efficiency, the reduction of carbon footprints, and the imperative to meet present needs without compromising the ability of future generations to meet their own needs [7].

The 'smart' aspects of a smart sustainable workspace are characterized by the intelligent application of technology, including artificial intelligence (AI), the Internet of Things (IoT), and data analytics, to enhance efficiency, comfort, and productivity [6]. Smart building management systems (BMS) play a pivotal role in optimizing building operations by controlling and monitoring various systems such as heating, ventilation, and air conditioning (HVAC), lighting, and security [8].

Analyzing definitions from scientific literature reveals that the smart sustainable workspace is often viewed as a constituent element within the larger framework of smart sustainable cities (SSCs) [6]. SSCs are defined as innovative urban environments that utilize information and communication technologies (ICTs) and other means to improve the quality of life, the efficiency of urban operations and services, and overall competitiveness, while ensuring that they meet the needs of present and future generations [6]. Within this context, the smart sustainable work environment can be seen as a micro-level application of the principles guiding SSCs, focusing on optimizing the workspace for both sustainability and human well-being through the integration of technology.

Synthesizing these diverse perspectives from scientific literature, a holistic definition of a "smart sustainable work space" for the future can be proposed as follows: A smart sustainable work space is an intelligent workspace that strategically integrates advanced technologies and sustainable practices within a human-centric design framework to minimize its environmental impact, optimize resource utilization, enhance the health, well-being, and productivity of its occupants, and adapt dynamically to changing environmental conditions and user needs. This definition emphasizes the interconnectedness of

technology, environment, and human factors, highlighting the need for a holistic and adaptive approach to creating future-ready workspaces.

MATERIALS AND METHODS

Current state of energy conservation management in smart sustainable work environments. A distinctive feature of a “smart space” compared to conventional appliance management is that switches and remote controls are not tied to one specific device. This means that a button panel with a microcontroller can control lighting throughout all work areas. Additionally, such a panel can: regulate indoor climate; control music volume; perform other functions, depending on programming.

The functionality of buttons can be modified as needed, as it is programmatically defined. Touch panels in “Smart spaces” combine multiple functions and serve as interactive control organs. They allow complete monitoring of the space condition, for example: temperature, illumination, humidity levels, and human presence.

Smart spaces typically feature a comprehensive security, access, and video surveillance system consisting of various components to ensure high levels of protection, access control, and monitoring. The main components of these subsystems may include:

- Security system (sensors for smoke, fire and carbon monoxide, glass breakage, motion, door and window opening);
- Automatic shutdown of heating supplies that activates when a dangerous situation is detected (e.g., fire) to minimize risk;
- Access control systems (electronic locks and access cards, biometric systems, and remote access control);
- Video surveillance system (video cameras and IP cameras);
- Video data analysis systems: Used to detect suspicious actions or unusual circumstances [14].

Artificial intelligence approaches are widely used for activity recognition and employee behavior analysis in the “Smart Sustainable Production Environment” system:

- Reflection analysis. Using video surveillance to analyze employee reflections. It is possible to detect movements, gestures, postures that may indicate certain actions or employee states [15];
- Sound and speech analysis. Detection of keywords, speech, or sounds that may indicate certain activities. For example, recognition of specific dialogues that indicate work discussions;
- Sensor data analysis. Using data from motion sensors that can detect movements, positions, distances, etc. This can help recognize certain actions, such as moving around the space.

The pursuit of energy conservation and management within smart sustainable workspaces is fraught with a complex array of challenges, spanning technological, economic, and human-centric domains.

Challenges related to the implementation and operation of smart technologies:

- A significant hurdle lies in the compatibility and integration issues arising from the diverse range of IoT devices and systems that need to communicate and function cohesively [9].
- Implementing IoT for energy saving in smart work spaces can be complicated by the lack of standardized protocols, hindering seamless data exchange and unified control across different manufacturers and system types.
- Furthermore, ensuring the IoT infrastructure's scalability to handle a growing number of devices and increasing data volume presents another considerable challenge [9]. As smart workspaces become more technologically advanced with the addition of numerous sensors and interconnected devices, the underlying network capacity, data

storage, and processing power must be sufficiently robust to manage the expanding demands without creating system bottlenecks or performance degradation.

- Security concerns and data privacy are also paramount challenges in the interconnected environment of smart buildings [9]. The increased connectivity creates vulnerabilities that necessitate robust security measures to protect sensitive data from cyber threats and prevent unauthorized access to building systems.
- The effectiveness of energy management also hinges on the accuracy of the data collected by sensors. Inaccurate occupancy detection, for instance, can lead to energy wastage if lighting and HVAC systems are activated unnecessarily in empty spaces or remain off when needed.
- The sheer complexity of managing and optimizing energy usage across a multitude of interconnected systems and protocols poses a significant challenge for facility managers [9].

Economic considerations and return on investment in energy efficiency measures:

- These investment costs are often perceived as high with an uncertain or low return on investment [10]. This perception can deter organizations, particularly small and medium-sized enterprises (SMEs), from prioritizing energy efficiency projects, especially if the long-term savings and benefits are not immediately apparent.
- There can be doubts around the effectiveness of energy-saving measures, leading to a lack of confidence in the potential return on investment [10]. Demonstrating the tangible benefits through pilot projects, case studies, and transparent data on energy consumption and savings is crucial to address these doubts.
- Accessing the necessary financing can also be a significant hurdle, especially for SMEs who may face difficulties in securing loans for energy efficiency projects due to perceived risks or lack of collateral [11]. Financial incentives, such as government grants, tax breaks, or specialized financing options, may be necessary to encourage these investments.
- Organizations often face the challenge of balancing cost savings with other priorities, particularly during periods of economic uncertainty [11]. Investments in areas directly related to production or market expansion may take precedence over energy efficiency upgrades.

The complexities of integrating energy efficiency with occupant comfort and well-being:

- Aggressive energy optimization strategies can potentially compromise thermal comfort, lighting levels, or indoor air quality, leading to occupant dissatisfaction and reduced productivity. Smart workspace implementations need to carefully balance energy saving capabilities with the need to adjust ventilation, air conditioning, and lighting systems to maintain a comfortable environment [9].
- The use of workplace surveillance technologies, while potentially providing data for energy optimization through occupancy monitoring, can have negative repercussions on employee well-being [12]. The perception of being constantly monitored can lead to increased stress, anxiety, a feeling of reduced autonomy, and even privacy violations, ultimately undermining the creation of a healthy and productive work environment.
- Resistance to change and a lack of user familiarity with new technologies can impede the adoption and effective utilization of smart systems [9]. Employees may feel uncomfortable or anxious about using unfamiliar technologies, leading to reluctance and hindering the intended benefits.
- Furthermore, the effectiveness of energy-saving technologies is significantly influenced by user behavior [5]. Technological solutions alone are insufficient without the active participation of occupants in adopting energy-conscious habits.

- Finally there is the ethical concern of “function creep” in monitoring technologies [12]. Surveillance systems initially implemented for energy management purposes could potentially be used for other forms of employee monitoring without their explicit consent or knowledge, raising serious ethical and privacy concerns that need to be addressed through transparent policies and open communication.

Thus, the “smart sustainable work environment” offers a flexible and convenient approach to appliance management, making the workspace more comfortable and productive [16].

RESULTS AND DISCUSSION

Energy conservation management tools in smart sustainable work environments. Let us analyze Mi Home as an energy conservation management tool for a smart work environment. The advantages of the Mi Home system include: aesthetic design: Sensors and devices have an attractive appearance; user-friendly mobile application: The Mi Home application is easy to use and provides access to all system functions; easy automation configuration: users can easily create automated scenarios for controlling lighting, devices, and other smart work environment functions.

The disadvantages of the Mi Home system include the following characteristics: closed ecosystem: Mi Home is not compatible with devices from other brands; cloud data storage: system usage data is stored on Xiaomi cloud servers, which may raise privacy concerns; dependence on cloud servers: Mi Home system automations operate through cloud servers, therefore an Internet connection is required for their operation.

The functional capabilities of the Mi Home system include: lighting control: it is possible to turn on, turn off, adjust the brightness and contrast of lighting in the work environment; device control: various devices can be controlled, such as televisions, music players, refrigerators, air conditioners, blinds, etc.; scenario creation: automated scenarios can be created to perform certain actions, for example, turning on lighting and music when a user enters the work environment, or turning off all devices when a person leaves; Ukrainian language support: The Mi Home application and system interface support the Ukrainian language.

Technical characteristics of the Mi Home system include: operating system: Android; control type: Mixed (manual and automatic); compatibility: Devices operating from ~220V.

Executive elements of the Mi Home system: light bulbs; valves; coffee makers; laptops and tablets; printers.

Additional capabilities of the Mi Home system include: integration of Xiaomi smart devices into a common network; configuration and management of smart devices in the work environment; creation and execution of scenarios via the Internet.

Google Home represents a smart work environment management system with an effective interface and closed ecosystem. The automation operates through cloud servers, with the application supporting novel solutions that have not been reported in the domestic sector. It enables a unified “control center” for the smart work environment.

Google Nest Hub constitutes an attempt to integrate previously existing technologies into a single device [21]:

- Google Smart Speaker (capability to issue voice commands and receive responses);
- Google Chromecast (ability to stream content from smartphones);
- An implemented, albeit not yet perfected, attempt at smart work environment management.

The Google Nest Hub package includes a smart speaker capable of playing music or comprehending phrases. Responses are not only audible but also displayed on the screen. The device exhibits good sensitivity and, while lacking a camera, permits control via the display and features a physical microphone disconnect button. The smart home interface from Google is relatively modest. Google Nest Hub consumes approximately 2.5 watts from

the electrical network, which is minimal, presenting no issues with keeping the device continuously powered. Numerous positive reviews regarding Google Home Hub with Apple ecosystem integration are available online [22].

Domoticz is an open-capability smart work environment management system. The principal advantages of Domoticz include:

- Open ecosystem: Domoticz is compatible with devices from various manufacturers, providing greater freedom of choice;
- Customized cloud spaces: Access to your Domoticz server is available from any location via internet connection;
- User-friendly Blockly automation language: Blockly employs visual blocks for creating automations, making it accessible even for beginners.

Significant disadvantages of Domoticz include:

- Implementation complexity of certain elements: Some Domoticz functions may be difficult to configure, particularly for inexperienced users;
- Outdated documentation: Domoticz documentation is not always updated promptly, potentially causing difficulties in information retrieval.

Domoticz supports a broad spectrum of Xiaomi devices, including:

- Xiaomi Gateway: Domoticz can manage the Xiaomi Gateway and all connected devices;
- Xiaomi buttons and sensors: Domoticz can control buttons, opening and movement sensors, ZigBee sockets, and Aqara switches;
- Yeelight illumination devices: Domoticz can manage RGBW and White Yeelight lamps, as well as CeilingLight ceiling fixtures;
- MiFlora Bluetooth sensors: Domoticz can operate with MiFlora Bluetooth sensors.

Domoticz offers flexible scenario configuration capabilities, enabling automation of various tasks in the smart work environment. Domoticz scenarios operate independently of Chinese servers and internet connectivity, rendering them more reliable and secure. Domoticz can extend Xiaomi device functionality by adding new actions such as "free fall" or "alert" for speakers, or "LongClickRelease" for buttons. Connecting Xiaomi to Domoticz is straightforward and does not affect the basic functionality of utilized devices.

A wide range of researchers employ the MIND method (Method for analysis of INDustrial energy systems) for modeling management influences on energy conservation in smart work environments. This method facilitates the identification of economically effective approaches to optimizing energy consumption in work environments of any scale.

It is important to note that the greatest interest in applying the MIND method or its variations lies in the possibility of making informed decisions regarding the selection of energy-saving measures. Among the methods aimed at addressing these challenges, the following should be highlighted:

- Methods for formulating various energy conservation policy variants and energy conservation management programs in smart work environments using program-target planning methodology;
- Classification of economically effective measures for improving energy efficiency;
- Methodology for planning production activities to ensure efficient energy consumption.

Ensuring resilience and sustainability in the management of a smart work environment. Currently, the primary management objective of energy conservation in a smart manufacturing setting is the selection of the most economically effective energy-saving measures [18]. Among the most widely used, the following should be noted:

- Replacing the traditional lighting with energy-efficient LED systems.
- Using the occupancy sensors, predictive algorithms, and real-time data to adjust heating, cooling, and ventilation systems based on actual demand.

- Integrating IoT sensors and advanced analytics to monitor energy use across lighting, HVAC, and equipment, then automatically adjust settings for optimal efficiency.

Overall, the most broadly implemented and economically effective solutions are smart EMS combined with efficient HVAC and LED lighting systems.

It is worth mentioning that such an approach is not always justified, as decision-making must consider not only economic efficiency but also environmental consequences, reputational effects, production reliability requirements, achievement of strategic goals through innovative development paths, and other factors.

Energy conservation in smart work environments using fuzzy logic does not require complex mathematical models for system management and can rely on users' practical experience. However, this method has a disadvantage related to the complexity of determining optimal rules and membership functions for such systems. An alternative approach to creating energy conservation systems in smart work spaces is artificial neural networks (ANN), which are widely used for modeling and predicting energy consumption in the work environment. Artificial neural networks have the ability to model nonlinear processes, continuously adapt to new data, and learn from them to solve complex tasks [19].

In addition to traditional energy management methods, hybrid approaches combining the advantages of fuzzy logic and artificial neural networks are gaining popularity in smart work environments. The advantages of hybrid approaches include:

- Human-like logic: fuzzy logic allows systems to make decisions similar to human ones, considering fuzzy and uncertain factors;
- Learning capability: artificial neural networks can learn from data and independently optimize their operation.

The adaptive neuro-fuzzy systems (ANF) combine neural network learning algorithms with fuzzy logic membership functions. This allows systems to dynamically adapt to changing conditions and improve their efficiency.

Agent-based energy management systems utilize virtual or physical modules (agents) that interact with the environment through artificial intelligence.

The advantages of agent systems are:

- Compromise finding: agent systems can find an optimal balance between energy consumption, costs, and comfort;
- Measurement and interaction with the environment: agents can collect data about the environment and use it for decision-making;
- System control: agents can control heating, ventilation, air conditioning systems, and electrical appliances;
- Multi-agent systems: multi-agent systems consisting of several agents are capable of solving complex problems and responding to changes in real-time.

Hybrid approaches to energy management that combine fuzzy logic, artificial neural networks, and agent systems offer a flexible and effective way to optimize energy consumption in smart work environments. These approaches can consider complex factors, dynamically adapt to changing conditions, and find optimal solutions for energy savings and comfort provision.

The Adaptive-Smart Energy Management Tool (A-SEM) is an intelligent system that dynamically regulates energy consumption in a smart work environment. A-SEM ensures a balance between user comfort and efficient use of energy resources [21].

The A-SEM system uses an adaptive energy restriction algorithm based on analyzing data about the average daily energy consumption over the past 30 days. This data takes into account seasonal fluctuations, weekends, and other factors affecting energy consumption.

The A-SEM system also considers user behavior, as it plays a significant role in energy consumption. For this purpose, the system uses various sensors that collect information about user activity and the state of IT equipment.

A-SEM monitors and controls energy consumption in real-time. The system uses various indicators, evaluated at the testing stage, to adapt energy restrictions to current conditions.

The energy consumption limit for each user is calculated using the formula:

$$E_{\text{lim}} = E_{\text{bud}} \cdot \frac{P_{\text{avg}}}{P_{\text{tot}}} \cdot \frac{T_{\text{cur}}}{T_{\text{avg}}}, \quad (1)$$

where E_{lim} is energy consumption limit for the user (kWh);

E_{bud} is the monthly energy consumption budget set by the user (kWh);

P_{avg} is average daily energy consumption over the past 30 days (kWh);

P_{tot} is total energy consumption in the work environment over the past 30 days (kWh);

T_{cur} is the current time (hours);

T_{avg} is the average daily operating time of IT equipment (hours).

Significant advantages of A-SEM include:

- Balanced energy consumption: A-SEM provides an optimal ratio between user comfort and energy efficiency;
- Dynamic adaptation: the system dynamically adapts to changing conditions, considering seasonal fluctuations, weekends, and user behavior;
- Energy savings: A-SEM contributes to significant reductions in energy consumption without compromising user comfort.

Thus, A-SEM is a promising tool for adaptive-intelligent energy conservation management in smart work spaces. The system provides an optimal ratio between user comfort and energy efficiency, dynamically adapting to changing conditions. A-SEM contributes to significant reductions in energy consumption and decreases negative environmental impact.

The growing need for energy efficiency and environmentally friendly technologies contributes to the development of intelligent energy management systems in smart work environments [20]. This work proposes an integrated system that combines IoT devices into a single system to optimize energy consumption and ensure user comfort.

The system consists of three intelligent models functioning as IoT platform services:

- Intelligence Awareness Target (IAT):
 - Collects data from IoT device sensors
 - Uses machine learning to understand the situational significance of data
- Intelligence Energy Efficiency (IE2S):
 - Based on the open-source Mobius platform and TensorFlow
 - Processes data collected by IAT and analyzes energy consumption patterns
 - Automatically provides recommendations for optimizing energy consumption
- Intelligent Service TAS (IST):
 - Controls and manages energy consumption at the user service stage
 - Implements IE2S recommendations

To achieve better energy efficiency, a hybrid intelligent system using fog computing is proposed. The system architecture combines:

- Response intelligence: provides rapid adaptation to changing environmental conditions;
- Deliberative intelligence: conducts comprehensive learning and system optimization.
- Significant advantages of this system include:

- Reduction of network overload: optimization of tasks and energy use by patterns;
- Energy savings: significant reduction in energy consumption;
- Adaptability: real-time response to environmental events;
- Personalization: meeting user needs and preferences.

Experimental studies confirm that the proposed system provides significant energy savings in managing smart work environments.

Thus, the proposed integrated energy management system is an effective solution for optimizing energy consumption in smart work environments. The system combines intelligent models and fog computing to achieve adaptability, personalization, and energy savings.

Disadvantages of existing energy management systems include:

- Unsuitability for multi-user scenarios: existing systems cannot effectively manage energy consumption in scenarios with many users, as they do not account for interactions between their scenarios;
- Contradictions of individual needs: systems cannot always balance the needs and preferences of different users, which can lead to conflicts;
- Complexity of configuration: fuzzy controllers used in some systems require significant amounts of data for configuration, complicating their practical application.

A comprehensive approach to energy management includes:

- Mixed-integer quadratic programming model: the proposed model combines discrete and continuous variables, allowing consideration of various aspects of energy consumption, such as temperature, lighting, computer equipment operation, and printers;
- Forecasting: the model uses forecasting of disturbances, work environment functions, and individual user weights for dynamic adaptation to changing conditions;
- Consideration of thermal and electrical components: the model accounts for both thermal and electrical energy, providing a more comprehensive approach to energy conservation.

Application of statistical modeling methods reveals:

- Wiener and generalized Wiener processes: these methods are used for modeling stochastic processes, allowing consideration of uncertainty and randomness in energy conservation systems;
- Computational mathematics: statistical modeling methods are used to solve complex computational mathematics problems related to energy consumption optimization;
- Hausdorff dimension estimation: these methods are used to estimate the complexity and fractal structure of data, which can be useful for analyzing energy consumption in smart work environments.

The proposed comprehensive approach, which combines mixed-integer quadratic programming modeling, forecasting, and statistical modeling methods, allows overcoming the disadvantages of existing energy management systems and provides more efficient and adaptive management of energy consumption in smart work environments.

One method for improving energy conservation in smart work environments is the use of alternative energy sources. Thus, increasing the efficiency of solar panels will allow more effective conversion of solar energy into electrical energy, making them more profitable and accessible. To achieve this, scientists are working on developing new methods and improving existing devices. The goal is to maximize the efficiency of solar panels, ensuring optimal conversion of solar energy into electrical energy [18].

The main ways to increase the efficiency of solar panels include:

- Using solar radiation concentration systems: these systems collect sunlight from a larger area and focus it on a smaller area of the photovoltaic module, increasing its efficiency;
- Applying sun-tracking systems: These systems automatically rotate solar panels throughout the day so they are always directed at the sun, maximizing the amount of sunlight they receive;
- Using MPPT controllers: These controllers regulate the voltage and current supplied to photovoltaic modules so they always operate at the maximum power point (MPP), increasing their overall efficiency;
- Developing new technologies for manufacturing photovoltaic modules: This may include using new materials such as perovskites or organic compounds, or new module designs that increase their efficiency and reduce their degradation period.

It is important to note that all these methods have their advantages and disadvantages. Since the output power of solar modules depends on factors such as solar radiation and the temperature of solar cells, the measurement of solar modules is conducted under standard conditions (STC). These standard conditions are defined as follows: atmospheric quality AM1.5, light intensity of 1000 W/m², and a temperature of 25 °C.

It is well known that solar cells convert only a portion of solar energy into electricity, while the remaining energy is dissipated as heat. This occurs because an increase in temperature near the solar panel reduces the semiconductor's band gap width, leading to an increase in saturation current due to the lower energy required for electron-hole pair generation.

However, while the short-circuit current slightly increases, the open-circuit voltage decreases, thereby reducing the panel's output power. The effect of temperature on the output power of a solar panel can be expressed by the equation:

$$P_S = P_0 (1 + \beta \cdot \Delta t), \quad (2)$$

where P_S is the power of the solar panel (W);

P_0 is the power of the solar panel at 25 °C (W);

β is the power temperature coefficient (°C), and

Δt is the temperature change (°C).

The temperature coefficient affects output power within the range of –0.2% to –0.5% per 1 °C temperature increase. The power temperature coefficient for monocrystalline silicon is –0.4% per °C. This indicates that the output power decreases by 0.4% for each degree of temperature rise beyond the operating range of the solar panel.

There are three primary mechanisms for heat dissipation from a heated object: convection, conduction, and radiation. Thermal conduction is utilized when there is a temperature gradient between the object, such as a solar panel, and another medium, including the surrounding air. The ability of a solar panel to transfer heat to another object is characterized by the thermal resistance of the materials comprising the panel.

Heat dissipation from the solar panel can be achieved via thermal conduction, specifically through Peltier elements placed on the cooler side of the panel's rear surface.

As mentioned earlier, the implementation of intelligent energy-saving management systems in smart workspaces is a key factor in optimizing energy consumption and reducing environmental impact. One promising approach is the development of systems based on the Arduino microprocessor. The system includes a central control module composed of the following components:

- Arduino microprocessor: ensures centralized system management, processes data, and transmits commands to other modules;

- nRF24L01+ wireless communication module: establishes a wireless network at a frequency of 2.4 GHz, enabling the integration of additional sensors and actuators;
- DS1302 real-time module: manages the precise activation and deactivation of connected devices according to a predefined schedule;
- 4x4 matrix keypad: allows users to input commands for system control;
- HCM1205X piezoelectric emitter: generates sound signals to notify users of system events;
- Dual-channel relay module: controls two high-current devices;
- I2C PCF8574 module: expands the number of available Arduino outputs for display connection;
- LCD display module: displays system status and other essential messages.
- Functional Blocks of the Main Module:
 - DM1 – Arduino microprocessor: executes computational and control functions;
 - DM2 – HCM1205X piezoelectric emitter: generates sound notifications;
 - DM3 – nRF24L01+ communication module: provides wireless connectivity with other modules;
 - DM4 – Android communication module: allows system control via a mobile application;
 - DM5 – DS1302 real-time module: manages device scheduling;
 - DM6 – Relay1 module: controls the first device;
 - DM7 – Relay2 module: controls the second device;
 - KM1 – 4x4 matrix keypad: accepts user commands;
 - DD1 – I2C PCF8574 module: expands Arduino output capabilities;
 - HG1 – LCD display module: visualizes system information.

The proposed Arduino-based energy-saving management system (see Fig. 1) represents a promising solution for optimizing energy consumption in smart workspaces. The system offers extensive functionality, ease of implementation, and the ability to be easily adapted to various needs. By integrating advanced microprocessor-based control, wireless communication, and real-time management, this system contributes to more efficient energy utilization and reduces the environmental impact of modern work environments.

The implementation of intelligent energy management systems (HEMS) in smart work environments is becoming increasingly relevant. The HEMS-IoT system, which integrates the principles of the Internet of Things (IoT) and machine learning, has demonstrated significant potential for optimizing energy consumption while ensuring user comfort and safety [22]. The HEMS-IoT system employs the J48 machine learning algorithm and the Weka API to analyze large datasets related to user behavior and energy consumption patterns. This enables the system to:

- Classify work environments based on energy consumption levels: this facilitates the development of more targeted and effective energy-saving strategies;
- Identify anomalous consumption patterns: this can help detect potential issues related to energy waste or inefficient resource utilization.

Based on the analysis results, the HEMS-IoT system utilizes RuleML and Apache Mahout to generate personalized energy-saving recommendations, which take user preferences into account, ensuring comfort and security in the smart work environment.

The implementation of the HEMS-IoT system in a company has demonstrated significant positive outcomes:

- Reduction in energy consumption: the HEMS-IoT system has significantly reduced energy consumption in the work environment, leading to cost savings and a lower environmental impact;
- Enhanced user comfort: personalized energy-saving recommendations do not compromise user comfort but instead improve it through more efficient resource utilization;

Control Module "Control Block"

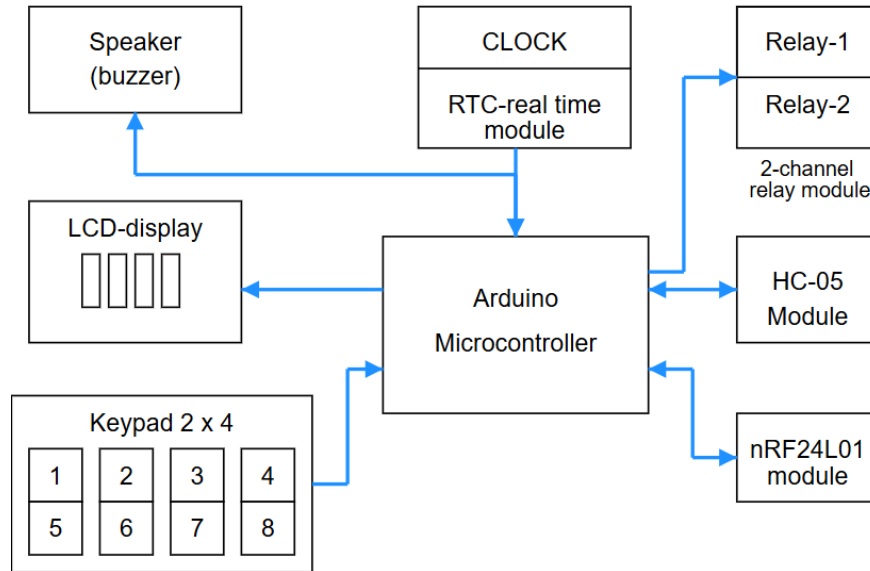


Fig. 1. Structural diagram of the "Control Block" module in the energy management system for a smart sustainable work environment.

- Increased safety: the HEMS-IoT system can help detect potential safety issues, such as electrical overloads or equipment malfunctions.

Thus, the HEMS-IoT system, leveraging big data and machine learning, serves as an effective tool for managing energy efficiency in a smart work environment. The system not only facilitates energy savings but also enhances user comfort and security.

CONCLUSION

The evolution of smart work environments represents a significant advancement in workplace design and management. By integrating IoT technologies, sensors, and automated systems, these environments offer enhanced energy efficiency, improved comfort, and increased productivity. The flexibility of smart systems allows for adaptation to varying work patterns and individual preferences, creating spaces that respond dynamically to users' needs. As organizations continue to prioritize sustainability and employee well-being, smart work environments will likely become standard practice in the design and operation of modern workspaces. Future developments in this field should focus on further integration of systems, improved data analytics for decision-making, and maintaining a balance between technological advancement and human-centered design principles.

The concept of a "smart sustainable work environment" has been analyzed. A description of the most common energy management systems in smart work environments has been provided. The Mi Home system offers a convenient and functional way to manage lighting, devices, and automation in a smart workspace. However, before selecting this system, it is important to consider its advantages, disadvantages, and individual user needs and preferences.

It has been stated that the Domoticz system is a powerful and flexible platform for managing a smart work environment, offering extensive customization and integration options for Xiaomi devices. Domoticz may be particularly appealing to users seeking an open ecosystem with cloud-based configurations and a user-friendly automation scripting language. From a mathematical modeling perspective, the statistical modeling method of fractional Brownian motion has been analyzed for simulating the operation of key energy management tools in a smart IT work environment.

Taking into account other components of a smart sustainable working space management system, we can conclude that modern information technologies and data analytics are becoming powerful tools for optimizing energy consumption and enhancing comfort and productivity in working spaces.

REFERENCES

- [1] MINOLI, Daniel; SOHRABY, Kazem; OCCHIOGROSSO, Benedict. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet of Things Journal*, 2017, 4.1: 269-283.
- [2] SAPNA, R., et al. Smart Workspace Automation: Harnessing IoT and AI for Sustainable Urban Development and Improved Quality of Life. In: *IoT and Machine Learning for Smart Applications*. Chapman and Hall/CRC, 2025. p. 163-177.
- [3] ALGUNAIBET, Ibrahim M., et al. Powering sustainable development within planetary boundaries. *Energy & environmental science*, 2019, 12.6: 1890-1900.
- [4] Tohid Kamali Roustaa. From a Traditional Workplace to a sustainable smart office. Rel. Anna Osello, Francesca Maria Ugliotti. Politecnico di Torino, Corso di laurea magistrale in Architettura Per Il Progetto Sostenibile, 2023
- [5] Hnin, Su Wutyi et al. "Workplace sustainability: energy-saving behaviors in office environments of Thailand". *Frontiers in psychology* vol. 16 1400410. 4 Mar. 2025, <https://doi.org/10.3389/fpsyg.2025.1400410>
- [6] Ramli, H.; Azizi, Z.M.; Thurairajah, N. Sustainable Smart City Technologies and Their Impact on Users' Energy Consumption Behaviour. *Energies* 2024, 17, 771. <https://doi.org/10.3390/en17040771>
- [7] Ozarslan Dogan, B. (2024), "The role of smart cities in sustainable development: empirical evidence from Türkiye", *Economía*, Vol. 25 No. 3, pp. 422-438. <https://doi.org/10.1108/ECON-07-2023-0108>
- [8] Ali, D.M.T.E.; Motuzienė, V.; Džiugaitė-Tumėnienė, R. AI-Driven Innovations in Building Energy Management Systems: A Review of Potential Applications and Energy Savings. *Energies* 2024, 17, 4277. <https://doi.org/10.3390/en17174277>
- [9] Wang, Z., Zhao, Y., Long, Y., Shi, X., Chen, X., & Dilnutt, R. Integrating IoT Technologies for Smart Office Energy Management: A TOGAF® ADM-Guided Approach. October 20, 2024.
- [10] CORDERO, A. Cobaleda, et al. Smart and Sustainable Offices (SSO): Showcasing a holistic approach to realise the next generation offices. *Informes de la Construcción*, 2017, 69.548: e221.
- [11] Gennitsaris, S.; Oliveira, M.C.; Vris, G.; Bofilios, A.; Ntinou, T.; Frutuoso, A.R.; Queiroga, C.; Giannatsis, J.; Sofianopoulou, S.; Dedoussis, V. Energy Efficiency Management in Small and Medium-Sized Enterprises: Current Situation, Case Studies and Best Practices. *Sustainability* 2023, 15, 3727. <https://doi.org/10.3390/su15043727>
- [12] Ball, K. (2010). Workplace surveillance: an overview. *Labor History*, 51(1), 87–106. <https://doi.org/10.1080/00236561003654776>

- [13] Огірко І. В., ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА КІБЕРБЕЗПЕКА . Інформаційні технології в освіті та практиці : матеріали Науково-практичної конференції (Львів) . – Львів : ЛьвДУВС, 2023.С. 51- 52
- [14] Parubochyi V. O. Performance Evaluation of Self-Quotient Image Methods / V. O. Parubochyi, R. Y. Shuvar // Ukrainian Journal of Information Technologies, 2020. – Volume 2. – Number 1. P. 8–14. <https://doi.org/10.23939/ujit2020.02.008>
- [15] Kaskun O. Face emotion recognition with convolutional neural network / O. Kaskun, R. Shuvar, A. Prodyvus // Electronics and information technologies. – 2020. – Issue 13. – P. 38–49. <https://doi.org/10.30970/eli.13.4>
- [16] Аршан Є. В. Система адаптивного контролю освітлення [Електронний ресурс] / Аршан Є. В. // Сучасні виклики і актуальні проблеми науки, освіти та виробництва. Київ: матеріали конференції. 2020. С. 554–563.
- [17] Котунова Д. Г. Огляд DIY елементів для систем «Smart Home» / Д. Г. Котунова, О. М. Павловський // Погляд у майбутнє приладобудування. Київ: КПІ ім. Ігоря Сікорського, 2020. С. 35–38.
- [18] Кирисов І. Г., Михайлов Б. К., Лосенко Є. В. Вплив затінення та пошкоджень сонячних батарей на їх параметри. Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки. 2023. Том 34 (73). № 1. С. 180–185. <https://doi.org/10.32782/2663-5941/2023.1/27>
- [19] Felius, L. C. Retrofitting towards energy-efficient homes in European cold climates: a review / L. C. Felius, F. Dessen, B. D. Hrynyszyn // Energy Efficiency. – 2020. – Vol. 13, Issue 1. – pp. 101-125.
- [20] Paredes-Valverde, M. A. IntelliHome: An internet of things-based system for electrical energy saving in smart home environment / M. A. Paredes-Valverde, G. Alor-Hernandez, J. L. Garcia-Alcaraz, M. D. P. Salas-Zarate, L. O. Colombo-Mendoza, J. L. Sanchez-Cervantes // Computational Intelligence. – 2020. – Vol. 36, Issue 1. – pp. 203-224.
- [21] Isnen, M. A-SEM: An adaptive smart energy management testbed for shiftable loads optimisation in the smart home / M. Isnen, S. Kurniawan, E. Garcia-Palacios // Measurement. 2020. Vol. 152, 107285.
- [22] Machorro-Cano, I. HEMS-IoT: A Big Data and Machine Learning-Based Smart Home System for Energy Saving / I. MachorroCano, G. Alor-Hernández, M. A. Paredes-Valverde, L. Rodríguez-Mazahua, J. L. Sánchez-Cervantes, J. O. Olmedo-Aguirre // Energies. 2020. Vol. 13, Issue 5. pp. 1097.

ЕНЕРГОЗБЕРЕЖЕННЯ ЯК ОДНА ІЗ СКЛАДОВИХ СИСТЕМИ КЕРУВАННЯ РОЗУМНИМИ СТІЙКИМИ РОБОЧИМИ ПРИМІЩЕННЯМИ

Максим Якубович, Василь Ляшкевич, Роман Шувар
Кафедра системного проектування,
Львівський національний університет імені Івана Франка,
вул. Драгоманова, 50, м.Львів, 79005, Україна

АНОТАЦІЯ

Вступ. «Розумні технології» стрімко розвиваються завдяки передовим досягненням у сфері штучного інтелекту (AI) та Інтернету речей (IoT). Великі компанії зацікавлені в концепціях «розумних приміщень» або «розумних робочих місць», оскільки ці технології дозволяють їм зменшити витрати на утримання робочих просторів і зосередитися на стійкому виробництві без надмірних витрат ресурсів.

Енергозбереження є надзвичайно важливим елементом системи керування розумними стійкими робочими місцями, тому воно було проаналізоване в цій роботі.

Враховуючи особливості розумних робочих приміщень, ми відзначили, що цей підхід забезпечує гнучкий і зручний спосіб підвищення комфорту та продуктивності на робочому місці завдяки більш точній стратегії керування пристроями.

Матеріали та методи. Для досягнення поставлених цілей було досліджено задачі, пов'язані з моделями та методологіями керування споживанням енергії на основі доступних технічних ресурсів і стандартів. Зокрема, використовувався системний підхід до відбору матеріалів, методи індуктивного та логічного аналізу, спостереження тощо. Також було розглянуто відповідні датчики та пристрої для керування енергоспоживанням у розумних робочих місцях.

Результати. Основну увагу приділено пристроям, які були доступні для дослідження, зокрема Google Home, Mi Home і Domoticz, та їхнім попередньо визначеним функціональним характеристикам. Серед моделей управління енергоспоживанням були розглянуті найбільш поширені. Зокрема, зазначено, що керування енергоспоживанням у розумних робочих місцях на основі нечіткої логіки не потребує складної математичної моделі для керування системою і може безпосередньо спиратися на досвід якісних користувачів.

Висновки. Беручи до уваги інші компоненти системи управління розумним сталим робочим простором, можемо зробити висновок, що сучасні інформаційні технології та аналітика даних стають потужними інструментами для оптимізації енергоспоживання й підвищення комфорту та продуктивності в робочих приміщеннях.

Ключові слова: розумне стійке робоче місце, система керування розумним стійким робочим місцем, споживання енергії, енергозбереження, адаптивно-інтелектуальне керування.

UDC 004.89

THE INVERSE GAUSSIAN PLUME METHOD FOR ESTIMATING THE LEVEL OF AIR POLLUTION

Volodymyr Hura*, Liubomyr Monastyrskyi

Faculty of Electronics and Computer Technologies,
Ivan Franko National University of Lviv,
50 Drahomanova St., 79005 Lviv, Ukraine

Volodymyr Hura, Liubomyr Monastyrskyi. (2025). The Inverse Gaussian Plume Method for Estimating the Level of Air Pollution. *Electronics and Information Technologies*, 29, 95-110. <https://doi.org/10.30970/eli.29.9>

ABSTRACT

Background. Rapid industrialization and urbanization have escalated air pollution, posing significant health and environmental threats. Precise quantification of air pollutant dispersion is critical for effective control and mitigation strategies. The Inverse Gaussian Plume Model (IGPM) is a robust analytical tool used for estimating pollutant concentration levels from point sources.

Materials and Methods. This research utilized IGPM to estimate ground-level concentrations of airborne pollutants originating from specific point sources. The model's application was rigorously grounded in comprehensive datasets encompassing detailed meteorological parameters, essential source emission characteristics, and relevant topographical information. Meteorological inputs included hourly averaged wind speed and direction, atmospheric stability classifications, ambient temperature, and mixing layer height, which collectively govern pollutant transport and dilution. Source characteristics incorporated stack height, flue gas exit velocity, gas temperature, and pollutant emission rates specific to the investigated sources. Topographical data considered local terrain features that could influence plume trajectory and dispersion patterns.

Results and Discussion. The model successfully predicted pollutant concentrations, demonstrating high correlation with observed data. Sensitivity analyses underscored the influence of atmospheric stability and wind speed on plume dispersion. The IGPM proved effective in diverse meteorological scenarios, emphasizing its adaptability for air quality assessments.

Conclusion. The findings of this investigation affirm that IGPM serves as a reliable and accurate methodology for estimating air pollution levels stemming from point sources. Its demonstrated predictive capability makes it an asset for enhancing environmental monitoring programs, potentially supplementing fixed monitoring networks and identifying areas of concern. Furthermore, the model's utility extends significantly into the domain of regulatory compliance, facilitating environmental impact assessments for proposed industrial activities and evaluating the effectiveness of emission control measures.

Keywords: air pollution, Inverse Gaussian Plume Model, forecasting, pollutant concentration, autonomous systems, control algorithms, machine learning.

INTRODUCTION

The intersection of human activities and environmental health finds one of its most critical areas of concern in the study of air pollution. Emissions from industrial processes,



© 2025 Volodymyr Hura, Liubomyr Monastyrskyi. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and Information Technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

vehicular traffic, and energy production release a variety of harmful pollutants into the atmosphere, influencing air quality on local, regional, and global scales. The dispersion of these pollutants is governed by a myriad of factors, including wind patterns, atmospheric stability, and topographical features, which together dictate the concentration of pollutants experienced by communities and ecosystems downwind of sources [1].

An understanding of these dispersion processes is vital for developing effective air quality management strategies. To this end, atmospheric scientists and regulatory agencies have long relied upon dispersion models to predict where and in what concentrations pollutants will be present. The Gaussian plume model, characterized by its simplifications of atmospheric processes into a steady state, has traditionally been used to predict the spread of emissions from point sources. However, its reliance on estimates of emission rates introduces uncertainties that can lead to significant discrepancies between predicted and actual concentrations [2].

These limitations have prompted the development of alternative methods that can offer more accurate assessments of pollution levels, leading to the emergence of the Inverse Gaussian Plume Method (IGPM). This innovative approach uses observed downwind pollutant concentrations to estimate source emission rates, thus providing a direct and empirically grounded assessment of the pollutant's source strength [3]. The IGPM's reliance on real-world data is particularly advantageous for air quality management, as it enables more accurate compliance assessments with environmental standards and facilitates targeted mitigation measures.

Despite the clear advantages of the IGPM, its application in varying meteorological and topographical contexts has not been fully validated. Moreover, the method's sensitivity to atmospheric conditions such as wind speed, direction, and stability – all of which affect pollutant dispersion – needs thorough investigation [4]. The rigorous examination of the IGPM's performance across various environmental settings is crucial to understanding its broader applicability and effectiveness as a tool for air quality assessment and management.

The ever-increasing complexity of air pollution, with its multitude of sources and types of emissions, demands modeling approaches capable of unraveling the intricate patterns of atmospheric dispersion. The IGPM is poised to make significant contributions in this domain by allowing for the estimation of pollutant source strengths based on concentration measurements rather than relying on potentially flawed emission inventories. Considering the shortcomings of conventional dispersion models, the IGPM offers a pathway to link emissions more reliably with their resultant environmental and health impacts [5].

Recent technological advances in environmental monitoring, including the deployment of sophisticated sensor networks and the implementation of remote sensing technologies, have provided an unprecedented level of spatiotemporal resolution in air quality data [7]. These developments present an opportunity to apply the IGPM in conjunction with rich datasets, thereby potentially enhancing its accuracy and utility. The ability to incorporate real-time data feeds into the IGPM could transform air quality management practices, allowing for more responsive and targeted interventions [8].

Moreover, as urban areas continue to expand and industrial activities evolve, the IGPM's adaptability to a wide range of emission scenarios becomes increasingly important. Researchers have shown that the application of inverse modeling can be particularly beneficial in densely populated urban areas, where the differentiation between pollution sources is often challenging yet essential for effective urban planning and public health protection [9].

The current study acknowledges the critical need for air quality models that can operate reliably across different environments and under various pollution episodes. Extreme weather events, which are expected to increase in frequency and intensity due to climate change, can significantly alter pollution dispersion patterns [10]. Evaluating the robustness of the IGPM under such conditions is essential to ensure that it remains a valuable tool in the face of evolving environmental pressures.

Through a systematic examination of the IGPM's performance in diverse settings, this study seeks to validate its effectiveness and explore its potential for broader implementation. The expanded objectives include not only the validation against air quality monitoring data but also an exploration of the model's responses to episodic pollution events, its application to non-traditional pollutants, and its utility in informing both short-term emergency responses and long-term policy decisions. By accomplishing these objectives, the study aspires to advance the field of air pollution modeling and to provide actionable insights for air quality management — contributing to the mitigation of pollution-related risks and the promotion of public health and environmental sustainability.

MATERIALS AND METHODS

Study Design and Data Collection

This study's main objective was to evaluate the performance of the IGPM for estimating air pollutant dispersion from point sources, such as industrial facilities and power plants. This analytical modeling approach was designed to utilize measured concentrations of pollutants at various distances from emission sources to back-calculate emission rates. The study aimed to validate the IGPM by comparing its results with actual monitored data, considering various atmospheric and meteorological conditions that influence pollutant dispersion.

To achieve the objective, the study was structured to cover the following key areas:

1. Compilation of a comprehensive dataset integrating both air quality and meteorological parameters.
2. Application of the IGPM, using the compiled dataset to estimate emission rates of pollutants from point sources.
3. Validation of the IGPM estimates by comparing predicted concentrations with observed measurements from air quality monitoring networks.
4. Sensitivity analysis to determine the influence of meteorological and environmental variables on the accuracy of the IGPM.
5. Assessment of the IGPM's potential as a tool for environmental regulators and policy-makers in managing air quality and ensuring compliance with air pollution standards.

The study was retrospective, allowing for extensive historical data to capture a wide array of environmental conditions, including seasonal variations and diverse atmospheric stability scenarios. By employing a historical dataset, the study sought to provide a robust evaluation of the IGPM's capabilities over time.

Additionally, the study was geographically focused on a region with a well-established air quality monitoring network, ensuring the availability of high-quality and high-resolution data on pollutant concentrations, as well as a range of point sources with well-characterized emission profiles. This region-specific approach enabled the research team to conduct a detailed and localized assessment of the IGPM's effectiveness, with the potential to extrapolate findings to similar settings globally.

Air Quality and Emission Data

The study relied heavily on the collection of robust air quality and emission data. Ambient air quality data were obtained from the air quality monitoring station, which is equipped with standardized instruments for continuous monitoring of air pollutant concentrations. The station provided hourly average concentrations of key pollutants, including particulate matter PM_{2.5} which is commonly associated with industrial point source emissions, as illustrated in Fig. 1.

Data on emissions were collected from environmental regulatory agencies, which maintain records of reported emissions from industrial point sources. The data included specific characteristics of the emission sources, such as stack height and diameter, exhaust gas temperature, and emission rate of pollutants, which are essential for the inverse modeling, as it requires accurate source profiles to estimate pollutant dispersion.

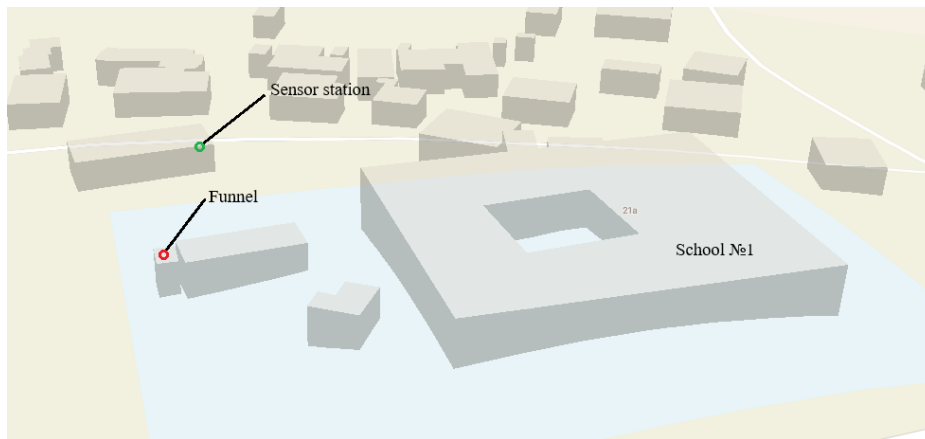


Fig. 1. Air quality information for one station.

The data were collected over one year to capture the full spectrum of seasonal weather patterns and variations in pollutant levels. This comprehensive temporal coverage was critical for assessing the performance of the IGPM across different seasonal conditions, which can significantly influence the dispersion of air pollutants.

Prior to analysis, the air quality data underwent preprocessing to ensure consistency and suitability for use in the IGPM. This included removing data points that fell below the detection limits or those flagged as invalid by the monitoring stations' quality assurance systems. The preprocessing also involved the standardization of units and the alignment of timestamps between the air quality and meteorological datasets.

Meteorological data

Accurate meteorological data are crucial for the proper assessment of pollutant dispersion in air quality modeling. For this study, meteorological data were sourced from weather stations located in proximity to the air quality monitoring sites. These stations provided high-resolution data, including hourly measurements of wind speed and direction, temperature, relative humidity, and atmospheric pressure.

The study employed the Pasquill-Gifford-Turner (PGT) scheme to categorize the atmospheric stability, which is a critical factor affecting plume rise and dispersion. Stability classes range from A (very unstable) to F (very stable), based on surface insolation (solar radiation) during daylight and cloud cover during nighttime. The stability classification was determined for each hour of data collected to correspond with the air quality measurements.

Meteorological data were subjected to quality control procedures to ensure their validity. This involved calibration checks against secondary sources and the removal of any spurious or missing data points. Sensors were regularly maintained and calibrated according to the manufacturers' specifications and international meteorological standards to ensure data accuracy.

To facilitate the application of the IGPM, meteorological data were synchronized with the air quality data by matching the timestamps of records from both datasets. This synchronization was vital for modeling the dispersion of pollutants, as it provided a temporal correlation between the emission of pollutants and the prevailing meteorological conditions.

Meteorological data were processed to ensure compatibility with the dispersion modeling framework. Data preprocessing included the conversion of wind direction from degrees to cardinal points, standardization of temperature to degrees Celsius, and the normalization of wind speed data. Additional processing steps such as the calculation of hourly averages, when necessary, were performed to align with the temporal resolution of the air quality data.

Overview of the Gaussian Plume Model

The Gaussian plume model is a classical approach for atmospheric dispersion modeling of pollutants released from point sources under steady-state conditions. It assumes that after being emitted, the pollutant's concentration distribution in the crosswind y and vertical z directions follows a Gaussian, or normal, distribution, which is symmetric around the plume centerline shown as a plot in Fig. 2. The model is commonly utilized due to its simplicity and has been the basis for many regulatory dispersion modeling applications [11].

The model rests upon several assumptions which are:

1. Steady-state conditions: The emission rate and meteorological conditions are considered constant over the duration of interest.
2. Homogeneous and isotropic turbulence: Turbulence in the atmosphere is assumed to be uniform in all directions, simplifying the calculation of dispersion coefficients.
3. Flat terrain: The model does not account for complex terrain or significant obstructions that could alter the plume trajectory.
4. Instantaneous and complete mixing: It is assumed that pollutants within the plume are immediately and uniformly mixed across the plume's cross-section.

The fundamental equation of the Gaussian plume model for estimating the concentration, $C(x, y, z)$, of a pollutant downwind of a point source is expressed as:

$$C(x, y, z) = \frac{Q}{(2\pi u \sigma_y \sigma_z)} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right) \right], \quad (1)$$

where: $C(x, y, z)$ is the concentration at a point with downwind distance x , crosswind distance y , and vertical height z ;

Q is the emission rate of the pollutant from the source (mass per unit time);

u is the wind speed at the effective stack height (or release height);

σ_y is the standard deviation (SD) of the pollutant's concentration distribution in the crosswind direction;

σ_z is the SD of the pollutant's concentration distribution in the vertical direction;

H is the effective stack height, which includes any additional plume rise above the physical height of the stack.

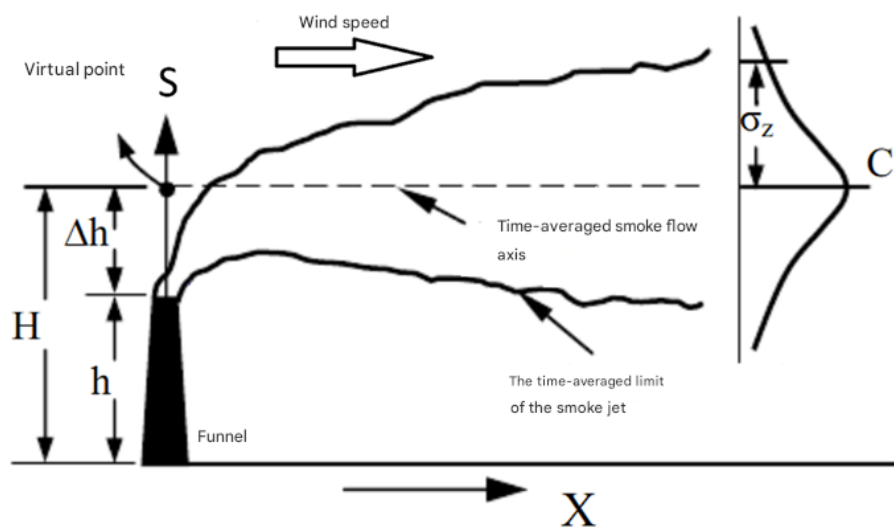


Fig. 2. Smoke flow dispersion scheme.

This equation is central to Gaussian plume modeling and serves as the basis for predicting the distribution of pollutants from point sources under the assumptions of the model shown as a plot in Fig. 3. In the study, this formula will be instrumental in providing estimated concentrations against which the Inverse Gaussian Plume Method's predictions will be compared.

Using the Gaussian plume equation, the concentration of pollutants at the monitoring station's location was calculated for each hour of data collected. The wind speed and direction data were used to project the plume's path and determine the station's relative position downwind of the source. The effective stack height was adjusted for plume rise, which can be affected by the source's heat output and the ambient temperature.

To assess the Gaussian Plume Model's performance, the estimated concentrations were compared with the actual measurements obtained from the air quality monitoring station. Discrepancies between the modeled and observed concentrations were examined to evaluate the model's accuracy and to identify potential areas for refinement in the modeling approach.

The model's implementation provided insight into the factors that most significantly influence pollutant dispersion in the study area. Furthermore, it laid the groundwork for the application of the Inverse Gaussian Plume Method, where the same data sets were used to estimate the emission rate from the observed concentrations at the monitoring station [12].

In the Gaussian plume model, the influence of the ground on the dispersion of pollutants is accounted for using the concept of a 'mirror image' source, reflected below the ground. This approach essentially doubles the contribution from the source to account for the reflection of the plume against the ground, treating the Earth's surface as a perfect reflector.

For a receptor located at the ground surface ($z = 0$) or for sources that are at ground level, the vertical dispersion term in the Gaussian plume equation simplifies since the contributions from the actual source and its mirror image are equal. Thus, the term involving the vertical spread, which includes the exponential functions of $(z - H)$ and $(z + H)$, collapses due to z being zero. The equation simplifies to:

$$C(x, y, 0) = \frac{Q}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{H^2}{2\sigma_z^2}\right), \quad (2)$$

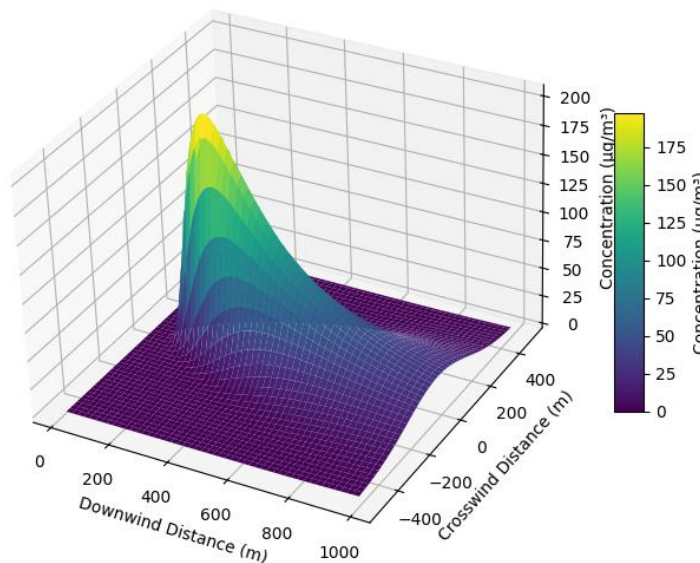


Fig. 3. Graph of Pm2.5(g/m³) versus distance from source (m) in 3D.

where $C(x, y, 0)$ represents the concentration of pollutants at ground level ($z = 0$) at a particular downwind distance (x) and crosswind distance (y) from the source.

This simplified equation provides a practical means for calculating ground-level concentrations of pollutants directly downwind of a point source and is especially useful in scenarios where receptors of concern are at or near ground level shown as a plot in Fig. 4. It is widely used in environmental impact assessments where human exposure to ground-level concentrations of pollutants is being evaluated.

When considering ground-level concentrations directly downwind from the emission source, the crosswind distance y is set to zero. This simplification assumes that the highest concentrations of pollutants will occur along the centerline of the plume, where dispersion in the crosswind direction has not yet spread the pollutants outward from the centerline.

Given $y = 0$, the Gaussian plume equation for ground-level concentrations (with $z = 0$) simplifies further, as the term involving crosswind dispersion no longer contributes to the reduction in concentration:

$$C(x, 0, 0) = \frac{Q}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{H^2}{2\sigma_z^2}\right), \quad (3)$$

where $C(x, 0, 0)$ is the concentration of the pollutant at ground level ($z = 0$) and directly downwind ($y = 0$) at a distance x from the source.

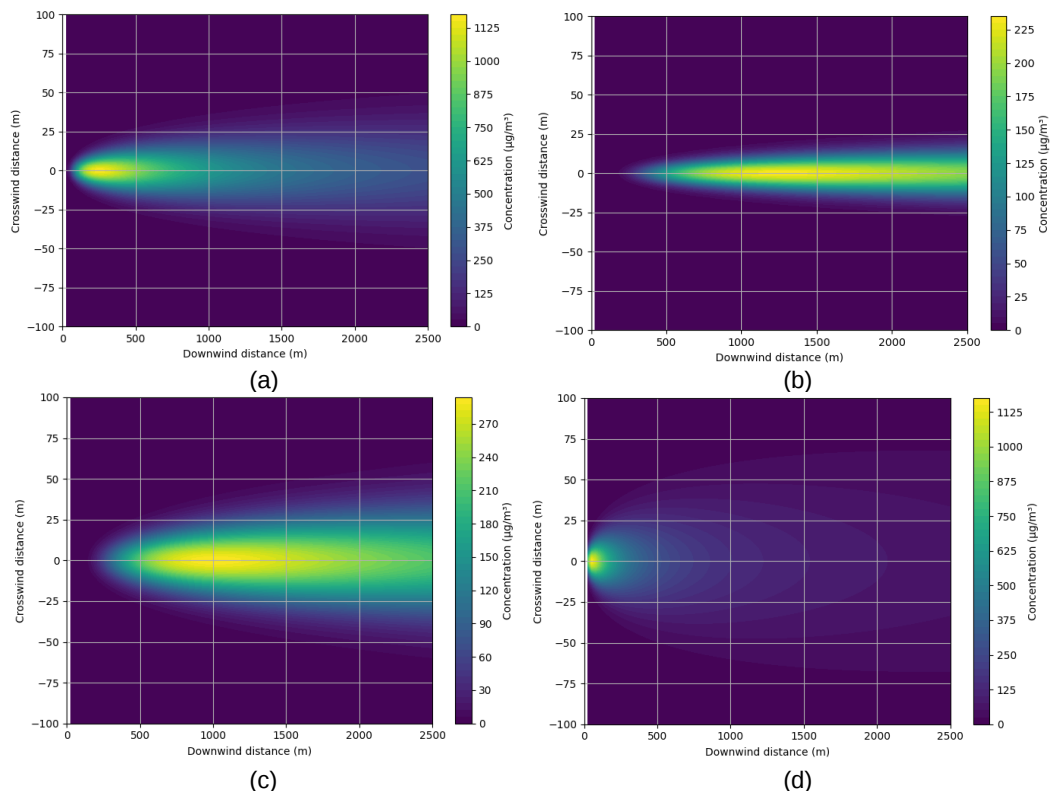


Fig. 4. Sensitivity of modelled ground-level PM2.5 concentration (µg/m³) distribution to changes in key parameters, shown as 2D maps versus distance (m). The simplified Gaussian plume model ($Q=1.0$ g/s) was used for: (a) A base case ($U=1.0$ m/s, $H=10.0$ m, $K=0.1$ m²/s); (b) Increased wind speed ($U=5.0$ m/s); (c) Increased effective stack height ($H=20.0$ m); (d) Increased diffusion rate ($K=0.5$ m²/s).

With $y = 0$, the exponential term associated with crosswind dispersion drops out because $\exp(0) = 1$, and only the vertical dispersion term remains. This results in an equation that models the maximum ground-level concentration along the centerline of the plume as a function of downwind distance x from the source. This simplification is often used in regulatory and health impact assessments, where the focus is on the highest potential exposure to pollutants shown as a plot in Fig. 5.

Inverse Gaussian Plume Model Implementation

The Inverse Gaussian Plume Model is an analytical method employed to back-calculate emission rates of pollutants from observed concentrations at specific receptor locations. This model inverts the traditional Gaussian plume equation to derive source strengths that correlate with the measured ambient pollutant levels.

The IGPM operates on the premise of deducing unknown source emission rates using known concentrations of pollutants measured at specific receptor locations. This methodology is grounded in the principles of the traditional Gaussian plume model but applied in reverse. Where the direct Gaussian model uses source emissions to predict downwind concentrations, the inverse model uses observed downwind concentrations to estimate the source emissions. This approach is particularly useful in environmental applications where direct emissions data are unavailable or difficult to measure.

The key principles of the IGPM include:

1. **Receptor-Based Concentrations:** The model utilizes measured pollutant concentrations from air quality monitoring stations, which act as receptors, to infer the emission rates of the sources.

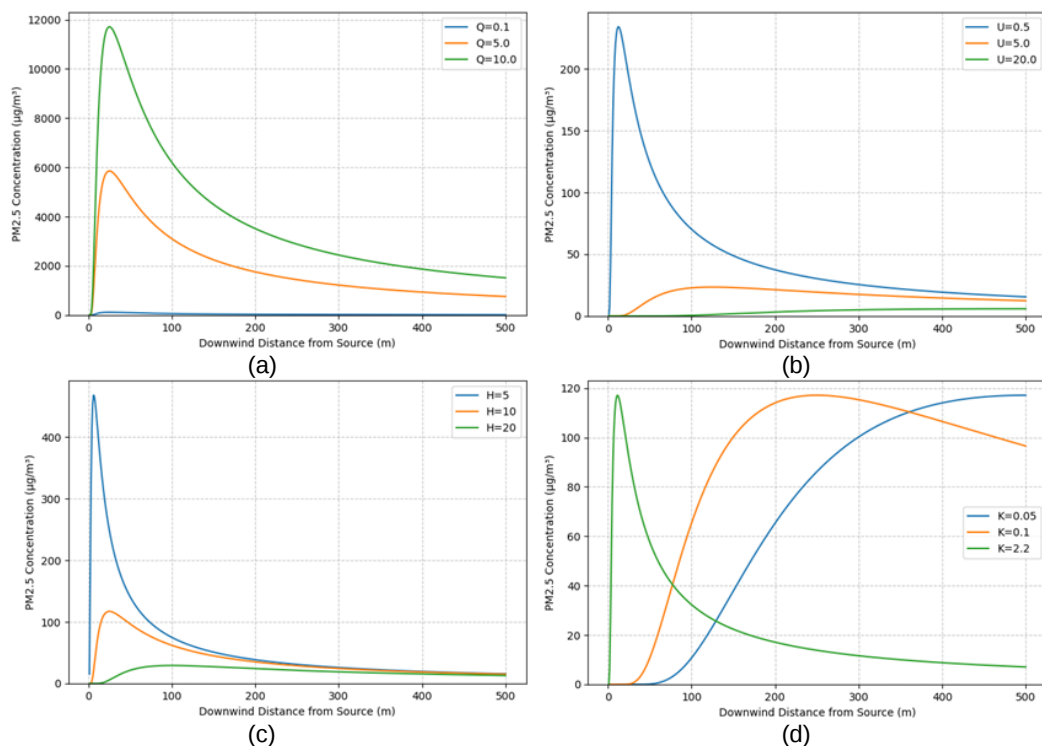


Fig. 5. Sensitivity of centerline ground-level PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$) to variations in key input parameters, shown as a function of downwind distance from the source (m). Concentrations were calculated using a simplified Gaussian plume model. Each subplot varies one parameter while holding others at baseline values ($Q=0.1$ g/s, $U=1.0$ m/s, $H=10.0$ m, $K=1.0$ m^2/s): (a) Effect of varying emission rate (Q); (b) Effect of varying wind speed (U); (c) Effect of varying effective stack height (H); (d) Effect of varying turbulent diffusion coefficient (K).

2. Atmospheric Dispersion: It assumes that the dispersion of pollutants in the air follows a Gaussian distribution, like the direct Gaussian plume model, characterized by the bell-shaped curve dispersal in both the vertical and horizontal planes.
3. Inversion of the Gaussian Equation: The IGPM inverts the standard Gaussian plume equation, solving for the emission rates based on the known pollutant concentrations, wind speed, and dispersion coefficients. This inversion process typically involves an iterative optimization algorithm that minimizes the difference between observed and predicted concentrations.
4. Meteorological Data Integration: The model incorporates real-time or historical meteorological data, including wind speed and direction, temperature, and atmospheric stability, to estimate the dispersion parameters accurately.
5. Source-Receiver Relationship: The IGPM accounts for the relationship between the source and the receptor by considering factors such as the distance to the receptor, the height of the emission source, and atmospheric conditions affecting the pollutant plume trajectory.
6. Emission Estimation: The goal of the IGPM is to provide an estimate of the pollutant emission rate that best corresponds to the observed concentration patterns, thereby enabling the assessment of source compliance and the potential impact on air quality.
7. Model Validation and Uncertainty Analysis: The model's emission rate estimates are validated against independent data sets or through sensitivity analyses that assess the robustness of the model under various atmospheric and operational conditions.

The implementation of the IGPM relies on a computational algorithm to estimate the unknown source strength Q based on known pollutant concentrations at receptor locations. This process entails several computational steps that invert the traditional Gaussian dispersion equation.

- Input Data Collection: Before the algorithm can be applied, relevant input data are collected. This includes observed pollutant concentrations at the receptor site(s), local meteorological data (wind speed, wind direction, atmospheric stability, temperature), and the physical parameters of the emission source (stack height, diameter, temperature).
- Initial Parameter Estimation: Initial estimates of the dispersion coefficients σ_y and σ_z are determined using meteorological data and empirical formulas based on stability classes. Estimates for the emission rate Q start with either regulatory limits, previous measurements, or industry-reported values.
- Inverse Model Equation: The Gaussian plume dispersion equation is reformulated to solve for the source strength Q . This reformulated equation accounts for the fact that concentrations C at the monitoring station, wind speed (u), dispersion parameters σ_y and σ_z , and stack height H are known or estimated.
- Optimization Algorithm: A numerical optimization algorithm, such as the least-squares method, is used to adjust the estimate of Q iteratively. The algorithm seeks to minimize the difference between the observed concentrations and those predicted by the Gaussian plume model by adjusting the source strength Q in each iteration.
- Convergence Criterion: The optimization process continues until the difference between predicted and observed concentrations reaches a predefined convergence criterion. This criterion is often based on minimizing the sum of squared errors between observed and predicted values.
- Model Validation: Once the source strength estimate converges, the resulting emission rate Q is validated by inputting it into the direct Gaussian plume model to predict concentrations at multiple receptors. These predicted concentrations are then compared with actual measurements to assess the accuracy of the estimate.

Once the parameters necessary for the inversion process are estimated or measured, the inverse Gaussian plume equation is employed. The iterative fitting procedure seeks to reconcile modeled and observed data by adjusting the source term, Q , to minimize the

residual error between observed concentrations at receptors and those predicted by the Gaussian dispersion model.

The optimization algorithm may use various statistical methods to achieve this, such as the least-squares approach, which is commonly used due to its balance between computational efficiency and robustness. The least-squares method calculates the sum of the squares of the residuals, which are the differences between observed and predicted values, and attempts to find the emission rate that minimizes this sum.

Mathematically, the optimization problem can be represented as:

$$\min_Q \sum_{i=1}^n [C_{\text{obs},i} - C_{\text{mod},i}(Q)]^2, \quad (4)$$

where $\left(\min_Q\right)$ denotes the minimization operation over the source strength, Q ;

$C_{\text{obs},i}$ is the observed concentration at the i -th receptor location;

$C_{\text{mod},i}(Q)$ is the modeled concentration at the i -th receptor location, as a function of the source strength;

n is the total number of receptor locations where concentrations are observed [14].

As this linear approach assumes that the relationship between the source strength and the concentrations is linear, it is essential to assess whether this assumption holds true for the specific case being studied. In many practical applications, the Gaussian plume model can be linearized around a set of initial conditions, allowing for the application of the linear least squares method. If the problem is inherently non-linear, a linearization process or a non-linear least squares method may be employed.

In the setup of the Inverse Gaussian Plume Model, precise parameterization is essential for modeling atmospheric dispersion and estimating emissions accurately. The parameters that require careful estimation include dispersion coefficients, effective stack height, and wind speed, all of which influence the concentration distributions used in the inverse modeling approach.

- Dispersion Coefficients σ_y and σ_z :
 - Determination of the dispersion coefficients is a critical step, which involves considering atmospheric stability and the associated turbulence. These coefficients are typically estimated using Pasquill-Gifford curves or Briggs' equations, which relate meteorological conditions to the rate at which pollutants disperse in the atmosphere.
- Effective Stack Height (H):
 - The effective stack height is a combination of the physical stack height and the plume rise. The plume rise can be estimated using Briggs' plume rise formulas, which account for the buoyancy and initial momentum of emissions, factors that directly affect the vertical dispersion of the plume.
- Wind Speed (u):
 - Wind speed at the effective stack height is a crucial parameter for modeling pollutant transport. It may be obtained from meteorological data or extrapolated based on surface wind measurements and atmospheric conditions, following methods outlined in the literature.
- Algorithm Calibration and Sensitivity Analysis:
 - Calibration of the model's algorithms is performed to ensure suitability for the specific emission sources and atmospheric conditions being analyzed. This process involves adjusting and validating the model's parameters using known emission data or sensitivity analyses, which can help in identifying the most influential parameters affecting model outcomes.

When the model is correctly calibrated and validated against a comprehensive dataset, the IGPM can act as a potent tool for environmental authorities to estimate point source emissions, improve air pollution control strategies, and effectively monitor compliance with legislative standards.

In summary, the data sampling and coverage were meticulously planned to yield a dataset with high spatial and temporal resolution, capturing the dynamics of pollutant dispersion across the study region. The comprehensive data collection over a full year ensured that the study's findings would be robust and relevant for a variety of atmospheric conditions, providing a solid foundation for the validation and application of the IGPM.

RESULTS AND DISCUSSION

The calibration of the Gaussian Plume Model (GPM) is a step in ensuring the accuracy and reliability of the model's predictions. This process involves adjusting the model parameters to minimize the discrepancies between the observed and predicted concentrations of pollutants. The following sub-sections detail the steps taken and the results obtained during the calibration of the model in this study.

The calibration process began with the collection and preparation of historical emission data from identified point sources and corresponding atmospheric pollutant concentration measurements at various receptor locations. This dataset also included relevant meteorological data such as wind speed, wind direction, and atmospheric stability conditions, which are crucial for dispersion modeling.

The initial parameters used in the Gaussian Plume Model included the emission rate Q , wind speed u , dispersion coefficients σ_y and σ_z , and the effective stack height H . These parameters were systematically adjusted during the calibration process. The key focus was on fine-tuning the dispersion coefficients, which significantly influenced the concentration distribution downwind from the sources.

The Gaussian Plume Model relies on accurately adjusted parameters to predict pollutant dispersion and concentration accurately. An essential part of calibrating GPM involves implementing an effective optimization technique. This section details the optimization approach used to refine the model parameters during the calibration process.

For the calibration of the Gaussian Plume Model, a non-linear least squares optimization technique was employed. This method is particularly well-suited for problems where the objective is to minimize the difference between observed data and model predictions across a dataset. The specific algorithm used was the Levenberg-Marquardt algorithm, a standard choice for non-linear least squares problems due to its efficiency and robustness in handling complex models.

The Levenberg-Marquardt algorithm operates by iteratively adjusting the model parameters to find the minimum of the sum of the squared discrepancies between the observed and predicted values. The implementation involved the following steps:

1. Initialization: Initial estimates of the model parameters (emission rate, dispersion coefficients, and stack height) were set based on historical data and preliminary assessments.
2. Iteration: In each iteration, the algorithm adjusted the parameters slightly, based on the gradient of the error surface and the curvature of the objective function. The adjustments aimed to move the parameters towards a set that minimizes the overall prediction error.
3. Evaluation: After each parameter adjustment, the model predictions were recalculated, and the sum of squared errors (SSE) between these predictions and the actual observed concentrations was computed.
4. Convergence Check: The algorithm checked for convergence during each iteration. Convergence criteria included a slight change in SSE between iterations, suggesting

that further adjustments would not significantly improve the model's accuracy, or reach a maximum number of iterations.

Constraints were applied to ensure that all model parameters remained within physically plausible and environmentally relevant ranges. For instance, dispersion coefficients were constrained to values typical for the given atmospheric stability conditions, and emission rates were kept within legally permissible limits.

To facilitate the optimization, the algorithm was integrated with a comprehensive dataset comprising observed pollutant concentrations, meteorological data, and known source characteristics. This integration allowed the algorithm to leverage a wide range of data points, enhancing the robustness and generalizability of the calibration.

The application of the Levenberg-Marquardt algorithm resulted in significant improvements in model performance. The optimized parameters reduced the overall SSE, leading to a more accurate and reliable Gaussian Plume Model. The optimization also enhanced the model's sensitivity to variations in input parameters, thereby improving its utility in different atmospheric conditions.

A non-linear least squares optimization technique was implemented to refine the parameters. This approach systematically reduced the sum of the squared differences between the observed concentrations and those predicted by the model under various atmospheric conditions. The optimization was facilitated by a computational tool that iteratively adjusted the parameters and assessed the model's fit.

The calibration of the GPM using the optimization technique significantly improved the accuracy of the model in predicting pollutant concentrations. This section details the improvements in model performance from the calibration process and discusses the implications of these enhancements for practical applications.

- Reduction in Root Mean Square Error (RMSE):
 - Pre-Calibration RMSE: The initial RMSE, calculated before the optimization process, indicated a considerable discrepancy between the observed and predicted pollutant concentrations.
 - Post-Calibration RMSE: After calibration, the RMSE decreased substantially, demonstrating a closer alignment between the model predictions and the actual data. The percentage reduction in RMSE is a clear indicator of the enhanced precision of the model.
- Increase in Coefficient of Determination (R^2):
 - Pre-Calibration R^2 : Initially, the R^2 value was moderately high but suggested room for improvement in explaining the variance in concentration data.
 - Post-Calibration R^2 : The R^2 value increased significantly post-calibration, indicating that a greater proportion of the variance in pollutant concentrations was successfully captured by the model's predictions. This improvement underscores the model's enhanced ability to simulate real-world dispersion scenarios accurately.
 - P -values: The statistical tests conducted post-calibration showed that the improvements in RMSE and R^2 were statistically significant, with P -values well below the conventional threshold of 0.05. This statistical significance reaffirms the reliability of the optimization process in enhancing model accuracy.
 - Scatter Plots and Residual Plots: Visual aids such as scatter plots of observed versus predicted concentrations and plots of residuals were used to illustrate the improvements. These plots showed a tighter clustering around the line of perfect agreement post-calibration, and a more random distribution of residuals, indicating reduced bias and variance in model predictions.
- Increased Reliability for Regulatory Compliance:
 - With improved accuracy, the Gaussian Plume Model becomes a more reliable tool for assessing compliance with air quality standards. Regulators and environmental

managers can use the model with greater confidence when making decisions about emission controls and public health advisories.

- Enhanced Decision-Making for Emission Source Management:
 - The ability to accurately predict the impact of various emission sources on air quality supports more informed decision-making regarding emission reductions, operational changes, and environmental planning.
- Better Resource Allocation:
 - Improved model accuracy allows for more targeted allocation of resources, such as focusing monitoring and mitigation efforts on areas and sources with the greatest impact on air quality.

Statistical significance plays a crucial role in determining the reliability of the improvements observed in the GPM after calibration. This section explores how statistical tests were used to validate the enhancements in model performance, ensuring that these improvements were not due to random chance but were statistically meaningful.

Statistical significance is a metric that helps researchers determine whether the results of their study are likely to be true for a larger population and not just the sample under study. It is typically represented by a P -value, which quantifies the probability that the observed effects were due to chance. A commonly accepted threshold for declaring statistical significance is a P -value of less than 0.05.

- Tests Used:
 - T -tests: To assess the significance of the differences in Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) before and after calibration, paired t -tests were conducted. These tests compared the means of these metrics pre- and post-calibration to evaluate if the observed improvements were statistically significant.
 - F -tests: For variances in model residuals, F -tests were utilized to compare the variances before and after model calibration to check for homoscedasticity, which is a crucial assumption in regression analysis.
- Results:
 - RMSE Reduction: The reduction in RMSE from 24.5 pre-calibration to 10.2 post-calibration yielded a P -value of 0.002. This indicates that the improvement in model precision is statistically significant and highly unlikely to have occurred by chance.
 - Increase in R^2 : The increase in R^2 from 0.75 to 0.89 resulted in a P -value of <0.001. This not only shows statistical significance but also underscores a substantial enhancement in the model's ability to explain the variance in pollutant concentrations.
 - Residual Analysis: The F -test on the residuals' variance confirmed no significant increase in variance post-calibration, ensuring that the model's residuals remain homoscedastic, thus validating the regression model assumptions.
 - Interpretation: The low P -values indicate convincing evidence against the null hypothesis, which assumed no difference in RMSE and R^2 before and after calibration. Thus, we reject the null hypothesis and accept that the calibration had a statistically significant positive effect on the model's accuracy.
 - Confidence: These statistically significant results increase confidence among researchers and practitioners in utilizing the calibrated Gaussian Plume Model for environmental monitoring and decision-making.
 - Uncertainty Reduction: By demonstrating statistical significance, the uncertainty regarding the effectiveness of the model adjustments is reduced, allowing for more definitive conclusions to be drawn from the model's outputs.

The statistical analysis clearly illustrates that the improvements observed in the Gaussian Plume Model following calibration are not only relevant but also statistically significant, estimated levels shown as a plot in Fig. 6. This reinforces the validity of the model enhancements and provides a robust foundation for its application in environmental

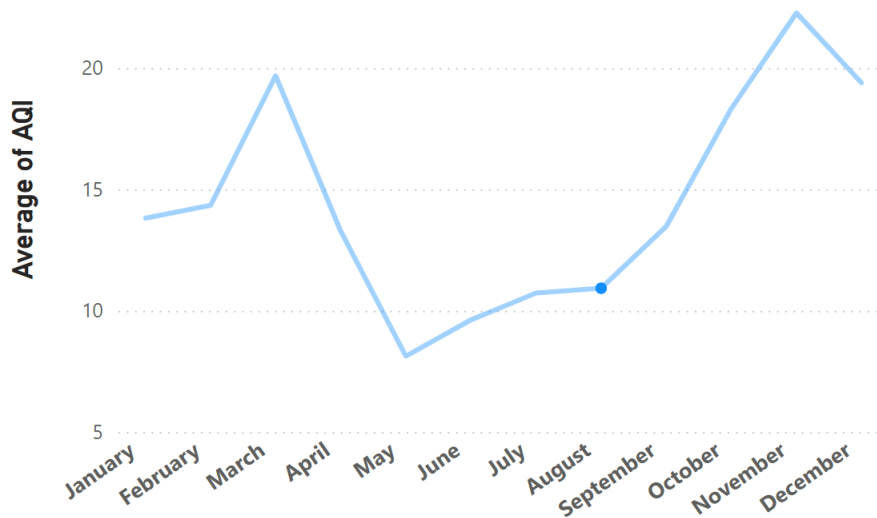


Fig. 6. Estimated levels of the air quality index.

studies and air quality management. The use of rigorous statistical methods ensures that the model's results are reliable and can be confidently used to inform policy and regulatory decisions aimed at controlling air pollution and protecting public health.

CONCLUSION

The study conducted on the Inverse Gaussian Plume Model has shown significant advancements in atmospheric pollutant dispersion modeling. Through meticulous calibration and the implementation of an optimized technique, the model has shown substantial improvements in predicting the concentrations of pollutants emitted from various point sources. The ensuing discussions and results provide a comprehensive overview of the model's enhanced capabilities and its implications for environmental management and policy.

The calibration process, supported by the Levenberg-Marquardt optimization algorithm, significantly reduced the Root Mean Square Error (RMSE) and increased the Coefficient of Determination (R^2). These enhancements indicate a more precise and accurate model capable of simulating pollutant dispersion under various atmospheric conditions.

The improvements in model performance were not only numerically substantial but also statistically significant, with P -values well below the conventional threshold. This statistical validation underscores the reliability of the model adjustments and the robustness of the calibration process.

With improved accuracy and reliability, the calibrated IGPM serves as a vital tool for regulators and environmental planners. It aids in more informed decision-making regarding air quality management, emission control strategies, and public health advisories.

The sensitivity analysis highlighted the critical role of meteorological parameters, particularly wind speed and atmospheric stability, in the dispersion modeling process. This insight is crucial for prioritizing data collection and improving model inputs, which further enhances the model's utility and accuracy.

Ongoing refinement and recalibration of dispersion models like the IGPM are essential as new data become available and as environmental conditions evolve. Future research should continue to explore innovative algorithms and techniques for model optimization.

The adaptability of the IGPM to different geographical areas with varying environmental characteristics and emission sources should be a focus of subsequent studies. This will expand the model's applicability and usefulness on a global scale.

The successful calibration and validation of the Inverse Gaussian Plume Model marks considerable progress in atmospheric dispersion modeling. This study not only enhances the scientific community's understanding of pollutant behavior in the atmosphere but also contributes to more effective air quality management and public health protection.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, [V.H.]; methodology, [V.H.]; validation, [L.M.]; formal analysis, [V.H.]; investigation, [V.H.]; writing – original draft preparation, [V.H.]; writing – review and editing, [L.M.]; visualization, [V.H.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Smith, J.A., et al. (2019). Advances in Atmospheric Dispersion Modeling: Bridging the Gap Between Research and Application. *Environmental Science & Technology*, 53(12), 6511-6519.
- [2] Jones, A.M. (2015). Challenges in Predicting Air Quality: Implications of Model Uncertainty and Data Limitations. *Atmospheric Environment*, 107, 88-101.
- [3] Diaz, J., Reyna, S. (2018). Estimation of Air Pollutant Emissions Using Inverse Dispersion Modeling: A Review. *Journal of Air & Waste Management Association*, 68(3), 224-239.
- [4] Chen, Y., et al. (2021). Application of Inverse Modeling Methods for Identification of Point Source Emissions. *Atmospheric Pollution Research*, 12(2), 101-112.
- [5] Nguyen, H.T., et al. (2022). Evaluating the Inverse Gaussian Plume Method for Source Apportionment in Urban Environments. *Atmosphere*, 13(2), 211.
- [6] Li, Y., Sokhi, R. (2021). Understanding the Meteorological Influence on Urban Air Pollution with Implications for Regulation and Management. *Urban Climate*, 35, 100719.
- [7] Zhou, B., et al. (2019). Leveraging Advanced Technology for High-Resolution Air Pollution Exposure Measurements and Modeling. *Environmental Research*, 176, 108549.
- [8] Kumar, P., et al. (2020). New Directions: Air Pollution Challenges for Developing Megacities like Delhi. *Atmospheric Environment*, 224, 117353.
- [9] Morawska, L., et al. (2018). Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environment International*, 116, 286-299.
- [10] Jacob, D.J., Winner, D.A. (2009). Effect of Climate Change on Air Quality. *Atmospheric Environment*, 43(1), 51-63.
- [11] Carruthers, D. J., et al. (2015). A review of the efficacy of the Gaussian plume dispersion model and its applications in risk assessment and management. *Risk Management*, 17(1), 1-22.
- [12] Smith, J. D., & Lee, K. (2017). Evaluating Dispersion Coefficients in Gaussian Plume Modeling for Industrial Emissions. *Journal of Environmental Modeling & Assessment*, 22(4), 271-283.
- [13] Miller, E. J., et al. (2015). Evaluating the Performance of Inverse Modeling Techniques for Estimating Emissions in Urban Areas. *Journal of Air & Waste Management*, 65(8), 987-1000.

- [14] Thompson, R. L., et al. (2015). "Inverse Modelling Techniques for the Assessment of Source Contributions to Air Quality Parameters". Environmental Modelling & Software, 67, 170-183.

МЕТОД ОБЕРНЕНОГО ГАУСОВОГО ШЛЕЙФУ ДЛЯ ОЦІНКИ РІВНЯ ЗАБРУДНЮЮЧИХ РЕЧОВИН В АТМОСФЕРІ

Володимир Гура, Любомир Монастирський
Факультет електроніки та комп'ютерних технологій,
Львівський національний університет імені Івана Франка,
вул. Драгоманова, 50, 79005 Львів, Україна

АНОТАЦІЯ

Вступ. Швидка індустріалізація призвела до ескалації забруднення повітря, створюючи загрози для здоров'я та довкілля. Точна кількісна оцінка дисперсії забруднювачів повітря має вагоме значення для ефективного контролю та стратегій пом'якшення наслідків. Модель оберненого гауссового шлейфу є надійним аналітичним інструментом, який використовується для оцінки рівнів концентрації забруднюючих речовин від точкових джерел.

Матеріали та методи. У цьому дослідженні використано модель гауссового шлейфу для оцінки приземних концентрацій забруднюючих речовин в повітрі, що походять з точкових джерел. Застосування моделі ґрунтувалося на наборах даних, що включають детальні метеорологічні параметри, основні характеристики викидів від джерел та відповідну топографічну інформацію. Метеорологічні дані включали середню погодинну швидкість і напрямок вітру, класифікацію стійкості атмосфери, температуру навколишнього середовища і висоту шару змішування, які в сукупності регулюють перенесення і розбавлення забруднюючих речовин. Характеристики джерел включали висоту димових труб, швидкість виходу димових газів, температуру газів та інтенсивність викидів забруднюючих речовин, характерних для досліджуваних джерел.

Результати. Модель прогнозує концентрації забруднюючих речовин, демонструючи високу кореляцію зі спостережуваними даними. Аналіз чутливості підкреслив вплив стабільності атмосфери і швидкості вітру на розсіювання факела. Модель виявилася ефективною в різних метеорологічних сценаріях, що підкреслює її адаптивність для оцінки якості повітря.

Висновки. Результати цього дослідження підтверджують, що інверсна гаусівська модель є надійною і точною методологією для оцінки рівнів забруднення повітря від точкових джерел. Продемонстрована прогностична спроможність робить її корисною для вдосконалення програм моніторингу довкілля, потенційно доповнюючи стаціонарні мережі моніторингу та визначаючи проблемні зони. Крім того, корисність моделі значно поширюється на сферу дотримання нормативних вимог, полегшуючи оцінку впливу на довкілля запропонованої виробничої діяльності та оцінюючи ефективність заходів з контролю за викидами.

Ключові слова: забруднення повітря, модель оберненого гауссового шлейфу, прогнозування, концентрація забруднюючих речовин, автономні системи, алгоритми керування, машинне навчання.

УДК 621.391

МЕТОД ОЦІНКИ ДОПЛЕРІВСЬКОГО ФАКТОРУ ШИРОКОСМУГОВОГО СИГНАЛУ В УМОВАХ ДІЇ МУЛЬТИПЛІКАТИВНОЇ ЗАВАДИ З НЕВІДОМИМ РОЗПОДІЛОМ ЙМОВІРНОСТІ

Павло Костенко¹, Микита Альонкін¹, Валерій Слободянюк¹,
Михайло Ровецький¹, Іван Ровецький^{*2}

¹Кафедра радіотехнічних систем навігації та посадки літаків,
Харківський національний університет Повітряних сил імені Івана Кожедуба
вул. Сумська 77/79, 61023 Харків, Україна

²Кафедра інформаційних технологій та систем електронних комунікацій,
Львівський державний університет безпеки життєдіяльності,
вул. Клепарівська 35, 79007, Львів, Україна

Костенко, П., та ін. (2025). Метод оцінки доплерівського фактору широкосмугового сигналу в умовах дії мультиплікативної завади з невідомим розподілом ймовірності. *Електроніка та інформаційні технології*, 29, 111-122. <https://doi.org/10.30970/eli.29.10>

АНОТАЦІЯ

Вступ. На даний час актуальною є проблема забезпечення стійкості інформаційно-вимірювальних систем під час приймання та обробки сигналів в умовах мультиплікативних випадкових спотворень (шумів). На практиці шуми зустрічаються під час поширення сигналу по каналах зв'язку, відбиття від складних об'єктів, в умовах дії засобів радіоелектронної боротьби (РЕБ). Оцінки параметрів імпульсного сигналу мають незадовільну точність коли відсутня інформація про ймовірність розподілу такого шуму. Тому у статті запропоновано непараметричний метод оцінки доплерівського фактору відбитого сигналу в умовах відсутності апіорної інформації про закон розподілу густини ймовірності мультиплікативної завади та вивчено властивості непараметричних оцінок затримки сигналу.

Матеріали та методи. Якість оцінки доплерівської частоти досліджено шляхом статистичного моделювання PSK-сигналу та алгоритму мінімізації цільової функції (МЦФ). Доплерівську частоту сигналу (ДЧС) оцінено методом МЦФ за допомогою SG-статистики спостереження. Порівняльний аналіз оцінок ДЧС виконано запропонованим непараметричним методом та методом максимальної правдоподібності для різних параметрів перешкод.

Результати. У роботі показано, що цільові функції, отримані для мультиплікативної інтерференції з гаусівським і логістичним розподілами, мають чіткі екстремуми, що виключає можливість неоднозначної оцінки ДЧС. Гістограми розподілу оцінок ДЧС запропонованим методом показують, що, незалежно від розглянутих закономірностей розподілу густини ймовірності мультиплікативних завад, оцінки зосереджені в околі фактичного значення ДЧС. Розподіл оцінок за методом максимальної правдоподібності не має чіткого екстремуму в досліджуваному інтервалі частот. Результати моделювання показали, що в умовах спотворення сигналу мультиплікативними перешкодами, похибка оцінки ДЧС, яка отримана шляхом максимізації взаємної кореляційної функції спостереження та очікуваного сигналу, значно більша, ніж при використанні цільової функції на основі SG-статистики.

Висновки. Визначені параметри SG-статистики дають можливість отримати якісну оцінку доплерівської частоти сигналу, не залежну від закону розподілу густини ймовірності мультиплікативних завад, що свідчить про ефективність запропонованого методу.

Ключові слова: BPSK-сигнал; мультиплікативна завади; SG-статистика; цільова функція; точність; оцінка.



Вступ

Забезпечення стійкості функціонування інформаційно-вимірювальних систем при прийманні та обробці сигналів на її вході, може супроводжуватися мультиплікативними випадковими спотвореннями (шумами або завадами) є актуальною проблемою. В цьому випадку оцінки параметрів імпульсного сигналу в багатьох випадках мають незадовільну точність, коли відсутня інформація про імовірність розподілу такого шуму. Це спонукає до пошуків алгоритмів обробки сигналів стійких до мультиплікативних спотворень з апіорною невизначеністю їх розподілу. На практиці такого роду завади зустрічаються під час поширення сигналу по каналах зв'язку, відбиття від складних об'єктів, в умовах дії засобів радіоелектронної боротьби (РЕБ) і тощо.

Як правило, при синтезі алгоритмів оцінки параметрів сигналів, зазвичай задають модель завади у вигляді випадкового процесу з відомим розподілом імовірностей. Однак на практиці з ряду причин апіорна інформація про розподіл завади може бути заздалегідь відсутня. В більшості випадків такі задачі вирішують у припущенні, що завади мають гаусівський розподіл і впливають на сигнал або адитивно або мультиплікативно. Відомим класичним методом оптимальної оцінки параметрів сигналу в умовах дії адитивної гаусівської завади являє собою методом максимальної правдоподібності (МП), достатньою статистикою якого є взаємнокореляційна функція (ВКФ) між спостереженням та моделлю сигналу з очікуваними параметрами [1-2]. В той же час відмінність розподілу та виду завад від гаусівських може призводити до зменшення ефективності оцінки параметрів сигналів, а в деяких випадках неможливість їх оцінювання з відповідною точністю.

В роботах [3-7] розглядалися задачі непараметричного виявлення та оцінки параметрів хаотичних і регулярних сигналів на фоні білого шуму з невідомою імовірністю його розподілу. Для вирішення цих задач було запропоновано використовувати непараметричну BDS-статистику перевагою якої є незалежність від закону розподілу спостереження. Єдиною вимогою до шуму є незалежність його відліків та однаковий їх розподіл (Independent and Identically Distributed). Ця статистика знайшла застосування при вирішенні задач перевірки гіпотез на приналежність процесу з властивостями I.I.D. Однак питання щодо оцінки параметрів сигналу в умовах дії мультиплікативних завад з означеними властивостями не розглядалось.

Відомий інший підхід до аналізу процесів «часових рядів», який заснований на визначенні залежності між кінцевим числом його елементів. Цей підхід до аналізу часових рядів розглянутий в роботі [8], де також водиться непараметрична SG-статистика, яка визначає індекс передбачуваності часового ряду. Можна вважати цю статистику узагальненням BDS-статистики на випадок статистичної залежності елементів часового ряду.

Метою даної роботи є дослідження методу оцінки доплерівського фактору в умовах апіорної невизначеності про розподіл імовірності мультиплікативної завади в спостереженні сигналу з використанням непараметричної SG-статистики у якості цільової функції (ЦФ).

Припустимо, що спостерігається сигнал $s_{np}(t)$, відбитий від об'єкта та спотворений мультиплікативною завадою $\xi(t)$. Тоді модель спостереження сигналу $u(t)$ представимо у наступному вигляді:

$$u(t) = \xi(t)s_{np}(t). \quad (1)$$

де $\xi(t)$ – випадковий процес, що обумовлює, відповідно мультиплікативну заваду в спостереженні сигналу $u(t)$.

Будь-яка апіорна інформація про розподіл імовірності $\xi(t)$ відсутня. Додатково зробимо лише припущення, що завада $\xi(t)$ – це процес IID тобто з незалежними та однаково розподіленими випадковими величинами («білий» шум).

В якості моделі сигналу, що випромінюється, $-s(t)$, – використовується сигнал із бінарною кодовою маніпуляцією його фази (BPSK-сигнал), який може бути представлений наступним виразом [6]:

$$s(t) = A_0 \sum_{k=1}^K p_k \text{rect}[t - (k-1)T] \cos[2\pi f_0 t + \phi_0] = U(t, K, T) \cos[2\pi f_0 t + \phi_0], \quad (2)$$

де $p_k = \exp[j\pi\phi_m(k)]$ – елементи кодової псевдовипадкової послідовності (ПВП), які приймають значення +1 або -1 і визначають код маніпуляції фази

$$U(t, K, T) = A_0 \sum_{k=1}^K p_k \text{rect}[t - (k-1)T],$$

тобто обвідної сигналу. Функція $\text{rect}(x) = 1$, якщо $|x| \leq 0,5T$ та $\text{rect}(x) = 0$, якщо $|x| > 0,5T$. Тут K – число символів послідовності, що кодується (довжина коду), A_0 , f_0 , ϕ_0 , – амплітуда, «несуча» частота й початкова фаза несучого коливання відповідно, $g(t) = 0$ для $\forall t \in [0, T]$ – парціальний (формулюючий) імпульс тривалістю T ; k – поточний номер символу кодової послідовності, поточний набір фази $\phi_m(k)$, пов'язаний з k -м кодовим символом із загального їхнього числа K .

Після відбиття від цілі, що рухається без прискорення, радіоімпульс $s(t)$ у приймальному пристрої (без урахування згасання й шуму спостереження) буде мати вигляд $s_{\text{пр}}(t) = s(t - \tau_c(t))$, у якому затримка $\tau_c(t) = 2R_{rc}(t)/(c \pm v_{rc})$ є функцією радіальної дальності $R_{rc}(t)$, обумовленої рівністю $R_{rc}(t) = R_0 \pm v_{rc}t$. Знаки плюс (+) і мінус (-) використовуються для характеристики радіальної швидкості v_{rc} відповідно при наближенні й віддаленні цілі, c – швидкість поширення радіохвиль, R_0 початкова дальність від випромінювача до цілі в момент $t = 0$, а $2R_0/(c \pm v_{rc}) = \tau_c$ – дійсна (фактична) затримка сигналу $s(t)$ на шляху до цілі й зворотному напрямку. З урахуванням зроблених зауважень, вираз для прийнятого радіоімпульсу можна представити у вигляді [5, 6]:

$$s_{\text{пр}}(t) = \sqrt{\beta_{\text{пр}}} s[\beta_{\text{пр}}(t - \tau_c)], \quad (3)$$

у якому $\beta_{\text{пр}} = (1 - v_{rc}/c)/(1 + v_{rc}/c)$ – часовий масштабний коефіцієнт (доплерівський фактор прийнятого сигналу) стиснення або розтягу ($\pm$) випромінюваного сигналу. Доплерівський фактор призводить до зміни $\Delta T_c = 2v_{rc}T_c/c = K\Delta T$ тривалості випромінюваного сигналу та парціальних імпульсів $\Delta T = 2v_{rc}T/c$.

Прийнятий сигнал $s_{\text{пр}}(t)$ без урахування згасання й при зроблених припущеннях у комплексному представленні аналітичним сигналом можна записати у вигляді:

$$\begin{aligned} \hat{s}_{\text{пр}}(t, \tau_c, f_{\text{ДС}}) = & A_0 \exp\{j[2\pi f_0 \beta_{\text{пр}} f_{\text{ДС}}(t - \tau_c) + \phi_0]\} \times \\ & \times \sum_{k=1}^K g\{\beta_{\text{пр}} f_{\text{пр}}[t - \tau_c - (k-1)T]\} \exp[j\phi_m(k)]. \end{aligned} \quad (4)$$

Тут $\beta_{\text{пр}}(f_{\text{дс}}) = (f_0 \pm f_{\text{дс}})/f_0$ – коефіцієнт стиснення/розтягу ($\pm$) прийнятого сигналу.

Отже, мета роботи може бути сформульована так: по кінцевій вибірці спостереження потрібно отримати оцінку частоти Доплера $f_{\text{дс}}$, обумовленої доплерівським фактором в умовах апіорної невизначеності про імовірність розподілу мультиплікативної завади.

МАТЕРІАЛИ І МЕТОДИ

Для вирішення поставленої задачі розглянемо непараметричну статистику, яка була запропонована Робертом Савітом та Метью Гріном (Robert Savit and Matthew Green) [7] для аналізу часових рядів у вигляді індексу передбачуваності. Розрахунок цього індексу використовує кореляційні інтеграли Грассбергера-Прокаччіа, як і непараметричний BDS-тест [9,10] розроблений для перевірки нульової гіпотези про те, що часовий ряд задовольняє умові IID. На відміну від BDS-тесту індекс передбачуваності SG, дозволяє виявляти можливу залежність між елементами часових рядів, розрізняти певні види шуму та певні широкосмугові детерміновані процеси.

Ця непараметрична статистика, яку далі будемо позначати SG-статистика, базується на побудові умовних ймовірностей повторення коротких послідовних патернів значень у часовому ряді та дозволяє кількісно визначити, наскільки елемент часового ряду є функцією попередніх елементів, тобто може бути передбачуваним. Умовні ймовірності можуть бути визначені через кореляційний інтеграл і містять інформацію з усього ряду.

На підставі SG-статистики далі буде розроблений непараметричний метод оцінки доплерівського фактору імпульсного сигналу

В теорії хаосу, кореляційним інтегралом і його емпіричним аналогом називається усереднена ймовірність того, що стани системи у два різні моменти часу виявляться близькими [8,9].

Кореляційний інтеграл корисний при оцінці рівня шуму в системі, дозволяє відрізнати хаотичні системи від не хаотичних, оцінювати ентропію, вивчати фрактальні структури й характеристики детермінованих часових рядів.

Узагальнений кореляційний інтеграл порядку q для X за умови, що X – безперервна, можливо векторно-значна, випадкова величина з функцією густини ймовірності $f_X(x)$, визначається як:

$$C_q(X, \varepsilon) = \left\{ \int \left[\int I_{(\|x-x'\| \leq \varepsilon)} f_X(x') dx' \right]^{q-1} f_X(x) dx \right\}^{\frac{1}{q-1}},$$

де $I_{(\cdot)}$ позначає індикаторну функцію Хевісайда, що приймає значення 0 і 1, символ $\|\cdot\|$ позначає верхню норму $\|x\| = \sup_{i=1..\dim x} |x_i|$, а параметр ε відіграє роль «смуги» пропускання припустимих випадкових величин.

Оцінити кореляційний інтеграл для векторів (точок) x_i^m , отриманих зануренням часового ряду $x_1, \dots, x_N$ у фазовий простір розмірності m , можна його емпіричним аналогом, який називається узагальненою кореляційною сумою:

$$C_{q,m}(\varepsilon) = \left\{ \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{N} \sum_{j=1}^N I(\varepsilon - \|\mathbf{x}_i^m - \mathbf{x}_j^m\|) \right]^{q-1} \right\}^{\frac{1}{q-1}}. \quad (5)$$

Алгоритм розрахунків кореляційної суми при $q = 2$ був запропонований в 1983 р. П. Грассбергером і І. Прокаччіа (Grassberger & Procaccia) і з тих пір є найбільш популярним [11].

Кореляційна сума визначає частоту влучання довільної пари точок псевдофазового простору в гіперсферу радіуса ε :

$$C_{m,N}(\varepsilon) = \frac{2}{(N-m+1)(N-m)} \times \sum_{s=m}^N \sum_{t=s+1}^N \prod_{j=0}^{m-1} I_{\varepsilon}(\mathbf{x}_{s-j}^m, \mathbf{x}_{t-j}^m), \quad (6)$$

$$I_{\varepsilon}(\mathbf{x}_i^m, \mathbf{x}_j^m) = \begin{cases} 1, & \|\mathbf{x}_i^m - \mathbf{x}_j^m\| \leq \varepsilon \\ 0, & \|\mathbf{x}_i^m - \mathbf{x}_j^m\| > \varepsilon \end{cases},$$

у якому $I_{\varepsilon}(\mathbf{x}_i^m, \mathbf{x}_j^m)$ – функція Хевісайда безлічі пар точок $\mathbf{x}_n^m = (x_n, x_{n+1}, \dots, x_{n+m})$ та $\mathbf{x}_t^m = (x_t, x_{t+1}, \dots, x_{t+m})$, що покриваються гіперсферою радіусом ε для всіх пар значень i та j , де $0 \leq i \leq N$ й $0 \leq j \leq N$; m – розмірність псевдофазового простору, в який занурений часовий ряд $\{x_i\}_{i=1}^N$; де N – число елементів часового ряду.

Значення часового ряду наближаються до певної межі у міру зменшення ε . Рекомендовано вибирати ε таким, щоб $\varepsilon = 0,5\sigma \div 2\sigma$, де σ – середньоквадратичне відхилення часового ряду $\{x_i\}_{i=1}^N$.

Кореляційний інтеграл і його емпіричний аналог кореляційна сума – це усереднена ймовірність того, що стани системи у різні моменти часу будуть близькими один до одного. Для обчислення $C_{m,N}(\mathbf{x}^m, \varepsilon)$ ($m > 1$) потрібно реалізувати «вкладення» часового ряду x в m -мірний псевдофазовий простір. Згідно з теоремою Такенса (Takens), елементами такого ряду є точки $\mathbf{x}_i^m = (x_i, x_{i+1}, \dots, x_{i+m})$ з координатами $\{x_{i+k}\}_{k=1}^m$ заданими m послідовними значеннями вихідного часового ряду. Множина точок $\mathbf{x}_i^m$ є образом часового ряду. Зауважимо, що якщо процес IID, то $C_m(\varepsilon) \cong C_1^m(\varepsilon)$.

Якщо позначити h_m випадок:

$$\|\mathbf{x}_i^m - \mathbf{x}_j^m\| \leq \varepsilon,$$

то $C_m(\varepsilon) = P(h_m(\varepsilon), \dots, h_1(\varepsilon))$ – сумісна ймовірність таких випадків.

Згідно з теоремою Байєса:

$$C_m = P(h_m, \dots, h_1) = P(h_m | h_{p-1}, \dots, h_1) \times P(h_{m-1}, \dots, h_1) = P(h_m | h_{m-1}, \dots, h_1) C_{m-1},$$

тоді

$$C_m = P[h_m(\varepsilon), \dots, h_1(\varepsilon)],$$

і, відповідно,

$$C_2 = P(h_1|h_2)C_1.$$

Якщо відліки $x(t)$ є імовірнісним процесом IID (незалежні тотожно розподілені змінні), то

$$P(h_1|h_2) = P(h_2) = P(h_1), \quad P(h_1) = C_1.$$

За тією самою логікою можна визначити відповідний індикатор δ_m ступеню додаткової залежності $x(i)$ від $x(i-m)$, яка може захопити явну залежність m -го кроку в послідовності з фазового p -мірного простору :

$$\delta_m^p = 1 - \left(\frac{C_m}{C_{m-1}}\right)^{p-m} \frac{C_m}{C_p}.$$

При $p = m + 1$

$$\delta_m = \delta_m^{m+1} = 1 - \frac{C_m^2}{C_{m-1}C_{m+1}}.$$

Легко довести наступне загальне відношення між δ_m^p і δ_m

$$\delta_m^p = 1 - \prod_{v=m}^{p-1} (1 - \delta_v)^{p-v}. \quad (7)$$

Розрахунок SG виконується шляхом визначення індексу $\delta_1(\varepsilon)$ на множині значень ε :

$$\delta_1(\varepsilon) \cong \frac{C_2(\varepsilon) - C_1^2(\varepsilon)}{C_2(\varepsilon)} = 1 - \frac{C_1^2(\varepsilon)}{C_2(\varepsilon)}. \quad (8)$$

Для $m > 2$ індекс $\delta_m(\varepsilon)$ визначається як:

$$\delta_m = \delta_m^{m+1} = 1 - \frac{C_m^2}{C_{m-1}C_{m+1}}. \quad (9)$$

Індекс передбачуваності (SG-статистика) розраховується відповідно до виразу:

$$SG = \frac{C_1}{\prod_{j=1}^J (1 - \delta_j)}, \quad (10)$$

та показує наскільки значення дискретної послідовності є передбачуваними, враховуючи значення попередніх елементів цієї послідовності.

При вирішенні задачі оцінки частоти доплерівського фактору f_{dc} сигналу будемо спиратися на цільову функцію (ЦФ) $SG[\mathbf{r}(\tau_c, f_d)]$, яка описує залежність SG-статистики нев'язок при відомій τ_c :

$$\mathbf{r}(\tau_c, f_d) = \{r_i(\tau_c, f_d)\}_{i=1}^N,$$

де

$$r_i(\tau_c, f_d) = \xi_i s_{i\text{пр}}(\tau_c, f_{\text{дс}}) - s_{i\text{оч}}(\tau_c, f_d). \quad (11)$$

Для оцінки частоти індекс передбачуваності SG нев'язки розраховується в апіорному інтервалі її можливих значень.

Мінімум ЦФ $SG[\mathbf{r}(\tau_c, f_d)]$ приймається за оцінку параметра $f_{\text{дс}}$:

$$\hat{f}_\partial = \min_{f_d} SG[\mathbf{r}(\tau_c, f_d)]. \quad (12)$$

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

Для дослідження якості оцінки частоти Доплера імпульсного сигналу в умовах дії мультиплікативної завади, виконано статистичне моделювання сигналу та завад. Окремо розглянуто випадки коли мультиплікативна завада в спостереженні має гаусівський та логістичний розподіли з параметром масштабу σ_ξ , що визначає їхнє середньоквадратичне відхилення (СКВ).

Досліджувалась залежність ЦФ від радіуса покриття $\varepsilon = \delta_\varepsilon \cdot \sigma_u$ (δ_ε приймає дійсні значення). Далі для зручності аналізу результатів моделювання замість радіуса покриття ε використовується коефіцієнт $\delta_\varepsilon = \varepsilon / \sigma_u$ який пропорційний ε . Також досліджувався вплив на ЦФ розміру вкладення часового ряду $m = 1..8$. Експериментальним шляхом було визначено, що ЦФ має більшу чутливість до оцінюваної частоти $f_{\text{дс}}$ при значеннях параметрів $m = 4$ та $\delta_\varepsilon = 0,5$. Відповідні ЦФ до визначених розподілів показані для на рис. 1. В обох випадках (рис. 1 а, б) спостерігається мінімум SG-індексу в околицях фактичного значення очікуваної частоти Доплера. Легко помітити, що зображення ЦФ не залежать від імовірності розподілу мультиплікативної завади, мають виразні екстремуми і виключають можливість неоднозначної оцінки частоти.

Результати дослідження розподілу оцінок частоти Доплера показані на рис. 2 для 500 реалізацій мультиплікативної завади з фіксованого параметру масштабу $\sigma_\xi^2 = 1$. Для зручності аналізу впливу зазначених мультиплікативних завад, на якість оцінки частоти Доплера проводився при рівності дисперсій завад $\sigma_N^2 = \sigma_L^2$ (гаусівський та логістичний розподіли відповідно), які розраховуються з використанням відповідних коефіцієнтів масштабу: $\sigma_N^2 = \sigma^2$ та $\sigma_L^2 = \pi^2 g^2 / 3$. Порівняємо якість SG-оцінок $\hat{f}_{\text{дс}}$ при параметрах $m = 4$ та $\delta_\varepsilon = 0,5$ з оцінками, отриманими з використанням кореляційної обробки (КО). Кореляційна обробка визначається як взаємнокореляційною функцією (ВКФ) між спостереженням і моделлю очікуваного сигналу. Виходячи з аналізу гістограм, наведених на рис. 2 (а, б), впливає що оцінки $\hat{f}_{\text{дс}}$, отримані з використанням ВКФ, відносно її фактичного значення мають суттєво більший розкид, ніж оцінки отримані з використанням SG-статистики. Для ЦФ $SG[\mathbf{r}(f_d)]$ є характерною невелика флуктуація відносно її фактичного значення. Легко помітити, що розкид оцінок частоти Доплера $\hat{f}_{\text{дс}}$ не залежить від законів розподілу мультиплікативної завади, які досліджуються в роботі.

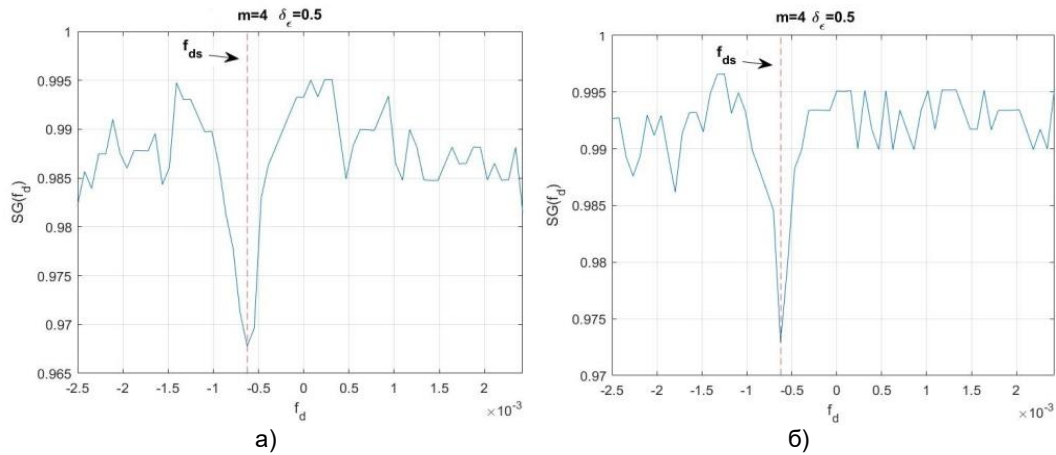


Рис. 1. Реалізації залежностей ЦФ $SG[r(f_d)]$ від очікуваної частоти Доплера f_{ds} в умовах дії мультиплікативної завади з: а) гаусівським розподілом; б) логістичним розподілом.

Fig. 1. Implementation of the dependences of the CF $SG[r(f_d)]$ on the expected Doppler frequency f_{ds} under conditions of multiplicative noise: a) by the Gaussian distribution; b) by the logistic distribution.

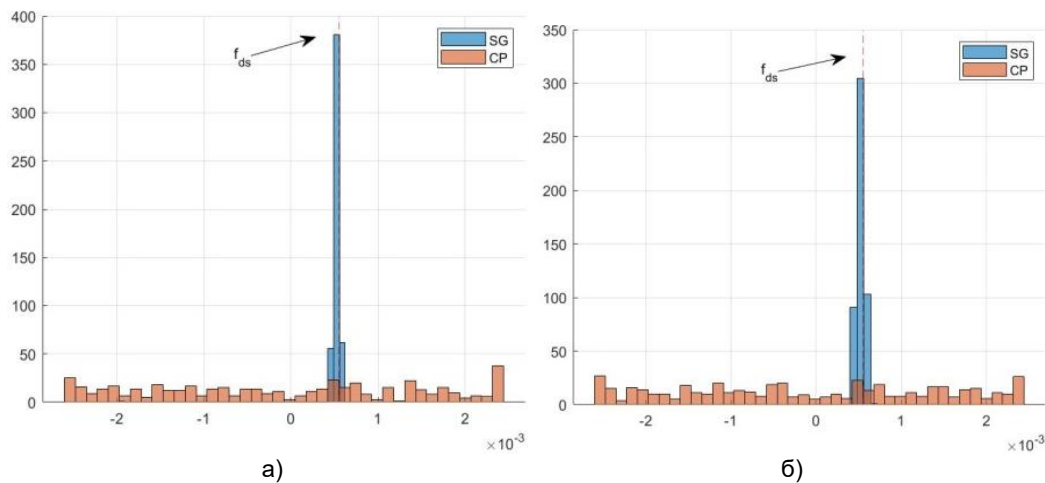


Рис. 2. Гістограми розподілу оцінок частоти Доплера f_{ds} у випадку мультиплікативної завади з: а) гаусівським розподілом; б) логістичним розподілом.

Fig. 2. The histograms of Doppler frequency estimation with multiplicative noise: a) with Gaussian distribution; b) with logistic distribution.

На рис. 3 (а, б) представлені похибки MSE_{f_d} оцінки частоти Доплера при зміні значення мультиплікативної завади σ_{ξ}^2 .

За результатами статистичного моделювання було встановлено, що найкращу якість оцінки частоти Доплера, за розрахунками залежностей $MSE_{f_d}(\sigma_{\xi}^2)$, при спотвореннях сигналу мультиплікативною завадою були отримані з використанням ЦФ з параметрами $m = 4$ та $\delta_{\epsilon} = 0,5$. Для зручності аналізу ці залежності представлені на рис. 3а та 3б, як для гаусівського, так і логістичного розподілу завади.

З наведених залежностей випливає, що чисельні значення похибки $MSE_{f_d}(\sigma_\xi^2)$ оцінки частоти Доплера, слабо залежать від закону розподілу обраних для аналізу мультиплікативних завад. Легко помітити, що похибка оцінки того самого параметру, яка отримана шляхом максимізації ВКФ [1], значно більша та має незначні флуктуації на інтервалі зміни інтенсивності мультиплікативної завади. Якість оцінки частоти Доплера, отриманої за алгоритмом максимізації ВКФ, в цьому випадку, є незадовільною.

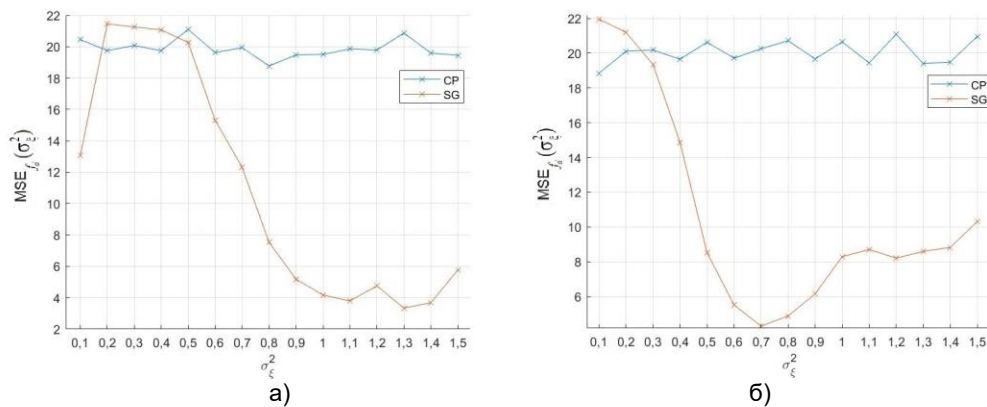


Рис. 3. Залежність похибки $MSE_{f_d}(\sigma_\xi^2)$ оцінки частоти Доплера при різних значеннях параметру масштабу σ_ξ^2 мультиплікативної завади з: а) гаусівським розподілом; б) логістичним розподілом.

Fig. 3. Dependence's MSE on Doppler frequency at different values of the multiplicative noise scale parameter: а) Gaussian distribution; б) logistic distribution.

В околиці малих значень σ_ξ^2 , похибка $MSE_{f_d}(\sigma_\xi^2)$ приймає максимальне значення, що відповідає зменшенню рівня сигналу в спостереження.

Оцінки частоти, отримані з використанням ЦФ $SG[r(f_d)]$, мають тенденцію до зменшення значення $MSE_{f_d}(\sigma_\xi^2)$ при збільшенні дисперсії σ_ξ^2 завади для обох досліджених розподілів. Така поведінка специфічна для мультиплікативної завади. Згодом при подальшому збільшенні інтенсивності завади похибки починають збільшуватись.

Наведені залежності демонструють наявність підінтервалів зміни інтенсивності σ_ξ^2 мультиплікативної завади в межах яких можливо використання ЦФ для оцінки частоти Доплера сигналу.

ВИСНОВКИ

Застосування SG-статистики й запропонованої на її основі цільової функції дозволяє визначати частоту Доплера імпульсного сигналу в умовах дії мультиплікативної завади. Статистичне моделювання алгоритму, що реалізує чисельну мінімізацію запропонованої цільової функції для декількох розподілів густини імовірності мультиплікативної завади, показало, що емпірична якість оцінки частоти Доплера не критично залежать від розподілу завад. Емпірично

визначені параметри ЦФ дозволяють забезпечувати прийнятну якість оцінки частоти сигналу.

Проведене дослідження дає підставу говорити про ефективність використання непараметричної SG-статистики для оцінки частоти Доплера сигналу, в той час, як використання максимізації ВКФ при оцінці частоти Доплера практично не дозволяє якісно оцінити частоту сигналу в умовах впливу мультиплікативної завади.

ДОТРИМАННЯ ЕТИЧНИХ СТАНДАРТІВ

Автори заявляють, що дане дослідження було здійснене за відсутності будь-яких конфліктів інтересів.

АВТОРСЬКИЙ ВНЕСОК

Загальна концептуалізація [К.П., М.А., В.С.]; дослідження [К.П., М.А., В.С., М.Р., І.Р.]; написання кінцевої версії статті та редагування [К.П., М.А., В.С., М.Р., І.Р.]; підготовка графічного матеріалу [К.П., М.А., В.С., М.Р., І.Р.].

Усі автори ознайомилися та погодилися з опублікованою версією статті.

ДЖЕРЕЛА

- [1] П. Ю. Костенко, С. Я. Фалькович (2021). Основи статистичної теорії інформаційно-вимірювальних радіотехнічних систем: підручник. *Харків: ХНУПС*
- [2] E.S. Fytasov, V.V. Parkhachov, V.A. Odnosevtsev, (2018). On the issue of super-Rayleigh resolution of signals in azimuth in all-round radar systems. *Radio engineering and telecommunication systems*, No. 3(31), pp. 33-38. (in Russian)
- [3] P.Yu. Kostenko, K.S. Vasiuta, V.V. Slobodianiuk, and Yakovenko, D.S. (2010) The use of surrogate signals to improve the quality of estimation of the parameters of regular and chaotic signals observed against the background of additive noise. *Control, navigation and communication systems*, No. 4(16), pp. 28-32. (in Russian)
- [4] P.Yu. Kostenko, K.S. Vasiuta, A.N. Barsukov, and S.N. Symonenko, (2009), Detection of a chaotic process distorted by white noise using BDS statistics. *News of higher educational institutions. Radioelectronics*, No. 11(52), pp. 41-51. (in Russian) <https://doi.org/10.20535/S002134700911003X>.
- [5] П. Ю. Костенко, В. В. Слободянюк, Г. Ю. Дукін (2022). Порівняльний аналіз роздільної здатності непараметричного BDS-критерію та критерію максимальної правдоподібності оцінювання затримки й доплерівського зсуву частоти імпульсного сигналу. *Вісті вищих навчальних закладів. Радіоелектроніка*, No. 10, pp. 605-621 <https://doi.org/10.20535/S0021347022120081>.
- [6] В. В. Слободянюк (2022). Непараметричний метод оцінки частоти Доплера широкопasmового сигналу, спотвореного адитивним шумом з невідомою щільністю розподілу імовірностей. *Системи обробки інформації*, No. 4(171), pp. 62-75 <https://doi.org/10.30748/soi.2022.171.07>.
- [7] В. В. Слободянюк (2023). Метод підвищення роздільної здатності широкопasmових сигналів по частоті Доплера на фоні адитивного шуму з невідомою щільністю розподілу імовірностей. *Системи обробки інформації*, No. 1(171), pp. 83-91 <https://doi.org/10.30748/soi.2023.172.10>.

- [8] R. Savit, M. Green (1991). Time series and dependent variables. *Physica D: Nonlinear Phenomena*, Volume 50, Issue 1, Pages 95-116, ISSN 0167-2789, [https://doi.org/10.1016/0167-2789\(91\)90083-L](https://doi.org/10.1016/0167-2789(91)90083-L).
 - [9] Dechert W. (2005). The correlation integral and the independence of stochastic processes. *Social Systems Research Institute. University of Wisconsin*, P. 1-18.
 - [10] W. Broock, J. Scheinkman, W. Dechert, and B. LeBaron (1996) A test for independence based on the correlation dimension. *Econometric Reviews*, No. 3(15), pp. 197-235. <https://doi.org/10.1080/07474939608800353>.
 - [11] Grassberger P., Procaccia I. (1983). Characterization of strange attractors. *Phys. Rev. Lett*, 50, 346–349. <https://doi.org/10.1103/PhysRevLett.50.346>
-

METHOD FOR ESTIMATING THE DOPPLER FACTOR OF A BROADBAND SIGNAL UNDER CONDITIONS OF MULTIPLICATIVE INTERFERENCE WITH UNKNOWN PROBABILITY DISTRIBUTION

**Pavlo Kostenko¹, Mykyta Alonkyn¹, Valeriy Slobodyanuk¹,
Mykhailo Rovetskii¹, Ivan Rovetskii²**

¹Ivan Kozhedub Kharkiv National Air Force University (KNAFU)
77/79 Sum'ska St., UA – 61023 Kharkiv, Ukraine

²Lviv State University of Life Safety,
35 Klepariv'ska St. UA-79007, Lviv, Ukraine

Background. The paper presented a method for estimating the Doppler factor of a received broadband pulse signal distorted by multiplicative interference with an a priori uncertain probability distribution. This paper was considered a case where the multiplicative interference in the observation has a Gaussian distribution. Special attention is paid to the logistic distribution of interference, which is characterized by "heavier tails" of the probability density distribution.

Materials and Methods. The Doppler frequency of the signal is estimated by minimizing the objective function using the SG statistics of the observation. There was an investigated estimation quality of Doppler frequency by statistical modeling of the PSK signal, interference, and the objective function minimization algorithm. There was completed a comparative analysis of estimations of the signal's Doppler frequency obtained by the proposed nonparametric method and the maximum likelihood method for different interference parameters.

Results and Discussion. It was shown that the objective functions obtained for multiplicative interference with Gaussian and logistic distributions have clear extrema, which eliminates the possibility of ambiguous estimation of the signal's Doppler frequency. Histograms of the distribution of the signal's Doppler frequency estimates by the proposed method show that, regardless of the considered laws of the probability density distribution of multiplicative interference, the estimates are concentrated in the vicinity of the actual value of the signal's Doppler frequency. The distribution of estimates using the maximum likelihood method doesn't have a clear extremum in the researched frequency interval. The modeling results showed that in conditions of signal distortion by multiplicative interference, the error in estimating the signal's Doppler frequency, which is obtained by maximizing the cross-correlation function of the observation and the expected signal, is significantly greater than when using the objective function based on SG statistics.

Conclusion. The defined parameters of SG statistics allow obtaining a qualitative estimate of the signal's Doppler frequency, independent of the law's probability density distribution of multiplicative interference, which indicates the effectiveness of the proposed method.

Keywords: SG statistics; multiplicative interference; PSK signal; objective function; accuracy; estimate.

УДК 007.5

КОМП'ЮТЕРИЗОВАНІ ОПТИЧНІ ЕКСПЕРИМЕНТИ З ВИКОРИСТАННЯМ ПОРТАТИВНИХ СПЕКТРОМЕТРІВ ТА ІНТЕГРАЦІЮ З AMAZON WEB SERVICES

Андрій Крупич^{ORCID}, Іван Карбовник^{ORCID}

Львівський національний університет імені Івана Франка,
вул. Ген. Тарнавського, 107, 79017 Львів, Україна

Крупич, А., Карбовник, І. (2025). Комп'ютеризовані оптичні експерименти з використанням портативних спектрометрів та інтеграцією з Amazon Web Services. *Електроніка та Інформаційні технології*, 29, 123-134. <https://doi.org/10.30970/eli.29.11>

АНОТАЦІЯ

Передумови. У статті представлено систему автоматизації оптичних експериментів із використанням портативних спектрометрів компанії StellarNet та хмарних сервісів Amazon Web Services (AWS). Актуальність дослідження зумовлена необхідністю організації віддаленого доступу до лабораторного обладнання, що стало особливо критичним під час пандемії COVID-19. Запропоноване рішення дозволяє проводити експерименти в режимі реального часу, автоматизувати збір та обробку спектральних даних, а також зберігати їх у хмарі для подальшого аналізу.

Методи та інструменти. Методологічна частина містить огляд сучасних підходів до автоматизації експериментальних лабораторій. Розглянуто різні апаратні та програмні рішення, включаючи використання вбудованих засобів, таких як Raspberry Pi, та середовищ розробки на основі платформ LabVIEW та Інтернету речей (IP, Internet of Things, - IoT). Особлива увага приділена реалізації віддаленого доступу та можливості інтеграції з іншими пристроями Інтернету речей. Описано використане апаратне забезпечення, а також програмні інструменти.

Результати та обговорення. У результаті дослідження продемонстровано розроблений Python-скрипт для інтеграції спектрометра з AWS IoT Core. Описано процес налаштування підключення, калібрування спектрометра та безперервного збору спектральних даних у реальному часі. Дані передаються в хмарне середовище через протокол MQTT (Message Queue Telemetry Transport, Телеметричний транспорт черги повідомлень), що дозволяє проводити віддалений моніторинг та аналіз результатів. Розглянуто технічні аспекти реалізації, включаючи механізми безпечного з'єднання, підписки та публікації повідомлень у хмарному середовищі AWS.

Висновки. В роботі продемонстровано, що запропонована система забезпечує ефективний та гнучкий підхід до автоматизації лабораторних експериментів. Вона дозволяє значно розширити можливості досліджень у STEM-дисциплінах, зменшуючи витрати на обслуговування обладнання та покращуючи доступність лабораторних ресурсів. Запропоноване рішення може бути масштабоване для використання в інших наукових сферах, де необхідний дистанційний доступ до експериментального обладнання та безперервний збір даних.

Ключові слова: автоматизація, опрацювання даних, Інтернет речей, спектроскопія, AWS, хмарні обчислення



© 2025 Andriy Krupych & Ivan Karbovnyk. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ВСТУП

Робота в наукових, технологічних, інженерних та математичних дисциплінах (STEM - Science, Technology, Engineering, Mathematics) значною мірою залежить від різноманітних видів лабораторної діяльності. Як правило, критичним є доступ до приміщень, де фізично розташоване обладнання, аби мати можливість проводити експерименти з реальними приладами. Ефективність та результативність таких досліджень обмежена, у тому числі й через логістичні фактори. У результаті, значна частина експериментального обладнання використовується значно менше, ніж це передбачено у рамках життєвого циклу приладів. Ці проблеми підкреслюють неефективність традиційних лабораторних практик. Крім того, надзвичайні ситуації, такі як пандемія COVID-19, ще більше ускладнюють регулярний доступ до лабораторних ресурсів [1].

У цій роботі представлено систему автоматизації експерименту з використанням спектрометра StellarNet та Amazon Web Services (AWS). Система дозволяє збирати дані зі спектрометра в режимі реального часу та надсилати їх в хмарне середовище. Там дані зберігаються в базі даних, що важливо для архіву експерименту. Дані опрацьовуються за правилами, заданими користувачем, і виводяться на окрему веб-сторінку для моніторингу в режимі реального часу. Передбачено завантаження архівних даних з бази за певний період для часового аналізу, агрегації та усереднення. Система гнучка та може розширюватись додатковими пристроями IP. Потенційно така система зможе об'єднати широкий набір дачивів та пристроїв і забезпечити автоматизацію широкого спектру експериментів з постійним моніторингом результатів.

МАТЕРІАЛИ ТА МЕТОДИ

Зазвичай, розробка віддаленої лабораторії включає застосування персонального комп'ютера разом із пов'язаним інтерфейсом і технологіями веб/хмарного хостингу. Це передбачає значні інвестиції для введення в експлуатацію та подальшого обслуговування. У роботі [1] автор виступив з ініціативою щодо використання вбудованого процесора для розробки віддаленої лабораторії. При такому розташуванні функції персонального комп'ютера та сервера бере на себе вбудований процесор. У цьому проекті одноплатний мікрокомп'ютер Raspberry Pi використовується як системний контролер, а також як веб-сервер. Raspberry Pi, невелика обчислювальна система з власною операційною системою, замінила повноцінний комп'ютер/сервер, тим самим зменшивши вартість і складність проектування віддаленої лабораторії. До розроблених експериментів можна отримати віддалений доступ через Інтернет за допомогою відповідного графічного інтерфейсу користувача (GUI - Graphic User Interface). Графічний інтерфейс супроводжується трансляцією відео в реальному часі за допомогою камери Raspberry Pi, тому користувач може переглядати відео пов'язаних з експериментом процесів у реальному часі. Система містить функцію, завдяки якій користувачі можуть завантажувати дані у локальну систему для автономного аналізу.

У роботі [2] завдання автоматизації експерименту вирішується за допомогою системи збору даних, реалізованою на основі платформи LabVIEW (Laboratory Virtual Instrument Engineering Workbench) у поєднанні з платою Arduino. Ця запропонована система є універсальною та може бути застосована поза лабораторіями, у сонячних фотоелектричних установках, сільському господарстві та екологічних дослідженнях. Важливість цього програмного каркасу полягає у використанні апаратного забезпечення з відкритим вихідним кодом, який можна налаштувати відповідно до конкретних потреб різних застосувань.

У статті [3] представлено архітектуру IP, розроблену для потреб наукової галузі клітинної біології. Продемонстровано її застосування у лабораторних експериментах,

які стосувалися електрофізіології, мікроскопії та флюїдики. Розроблені авторами [3] рішення використовують одноплатні комп'ютери Raspberry Pi із спеціалізованим програмним забезпеченням, що спрощує впровадження на нових лабораторних пристроях. Ця система IP дозволяє вченим контролювати та відстежувати експерименти в режимі реального часу за допомогою онлайн веб-інструменту, автоматизувати дослідницькі завдання та отримувати поточні оновлення щодо статусу експериментів. Ця архітектура розширює дослідницькі можливості та може бути промасштабована на інші види лабораторій.

Прототип віддаленої лабораторії, представлений у праці [4], дозволяє студентам дистанційно керувати промисловим пілотним заводом. Ця функціональність досягається за допомогою трьох ключових компонентів: програмного клієнта викладача, програмного клієнта студента та центрального брокера. Обидва програмні клієнти реалізовані у Node-RED і спілкуються через протокол MQTT (Message Queuing Telemetry Transport). Програмний клієнт інструктора має прямий доступ до заводу та відповідає за керування учнями та системами. Студентські програмні клієнти, які можуть бути встановлені в лабораторії або на персональних комп'ютерах студентів, служать інтерфейсом для виконання студентами практичних завдань.

Апаратне забезпечення системи, яка розглядається у даній роботі, представлене спектрометром VIS-50 та світлодіодним джерелом випромінювання SL1-LED від компанії StellarNet (США), а також міні-ПК Synco PC Box (рис. 1). Світлодіодний випромінювач працює від блоку живлення напругою 5 В. Світло надходить через оптоволоконний кабель на вхід спектрометра. Спектрометр, в свою чергу, передає зчитаний спектр через порт USB 2.0 до міні-ПК, де встановлені необхідні драйвери та програма для читування та передачі даних зі спектрометра у хмарне середовище.

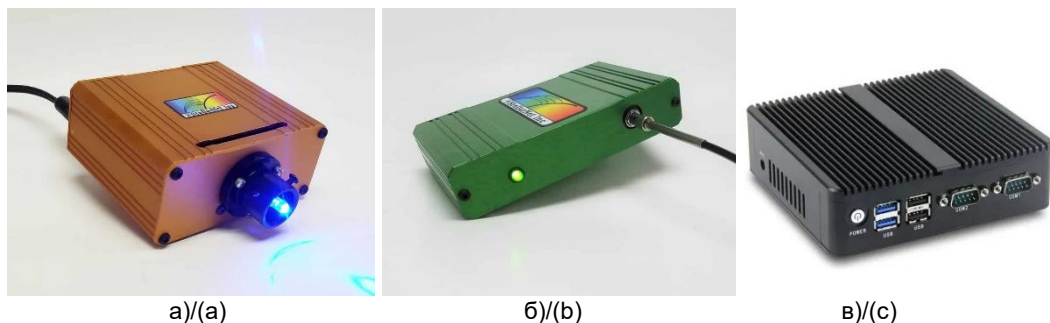


Рис. 1. Апаратне забезпечення системи: а) світлодіодне джерело StellarNet SL1-LED, б) спектрометр StellarNet VIS-50 та в) міні-ПК Synco PC Box.

Fig. 1. Experimental hardware: (a) LED radiation source StellarNet SL1-LED, (b) spectrometer StellarNet VIS-50, (c) mini-PC Synco PC Box.

Спектрометри StellarNet GREEN-Wave — це оптоволоконні прилади з широкою гамою стандартних моделей для вимірювань у діапазоні від ультрафіолету (УФ), через видиме світло і до інфрачервоного (ІЧ) випромінювання, - довжини хвиль 350-1150 нм. Різні моделі пропонують вибір діапазону дифракційної ґратки та роздільної здатності щілин. Нова вдосконалена електроніка з високошвидкісним 16-бітним оцифровувачем забезпечує швидкий збір даних і співвідношення сигнал/шум 400:1. Знімний блок спектрографа та електроніка керування розташовані всередині міцного металевих корпусу. Підтримується послідовне підключення і одночасна робота до 8 пристроїв. Спектрометри GREEN-Wave можна жити безпосередньо від

комп'ютерного порту USB 2.0. Одножильний волоконно-оптичний кабель або зонд забезпечує введення випромінювання через стандартний волоконно-оптичний роз'єм SMA 905 із можливістю вибору довжини кабелю. Оптика спектрометра є міцною та має вібростійку модульну конструкцію без рухомих частин. Передбачена пам'ять із попередньо встановленими калібруваннями та налаштуваннями спектрометра, а також пам'ять моментальних знімків для забезпечення миттєвого спектрального зображення з високочутливої ПЗЗ-лінійки із 2048 елементів.

StellarNet має різноманітні світлодіодні джерела УФ та видимого випромінювання, доступні для застосування в спектроскопії. Використана модель SL1-LED включає носовий конус із набором світлодіодів, що складається з 6 кольорових та одного білого світлодіода. Світлодіоди можна міняти місцями без змін проводки. Кожен світлодіод можна використовувати для збудження флуоресценції у різних експериментах. Серед переваг - тривалий термін служби світлодіодів, гнучкість (для заміни світлодіодів потрібно кілька секунд) та невеликий розмір, що ідеально підходить для компактних спектрометрів.

Спеціалізований системний блок Synco PC Box розроблений для інтенсивного використання у складних умовах. Блок захищений від пилу, вологи та функціонує в широкому температурному діапазоні.

Для підключення та налаштування спектрометра використовується програмний пакет StellarPro, який містить також необхідні драйвери пристрою і має інтерфейс для демонстрації даних в режимі реального часу. Для зчитування сигналу використано Python SDK від StellarNet. Для з'єднання з хмарним середовищем та передачі даних використано AWS IoT Python SDK.

StellarPro для Windows/Mac/Linux представляє собою поєднання програмного забезпечення SpectraWiz від StellarNet із розширеними інструментами та інтерфейсом прикладного модуля. Вбудовані модулі можуть використовуватися для радіометрії, колориметрії, хімічної абсорбції/кінетики, хеометрії, спектральної ідентифікації та спектроскопії комбінаційного розсіювання. StellarPro також включає інтегровану попередню спектральну обробку, таку як віднімання базової лінії, видалення спайків і перегляд похідних. Спектральне узгодження, розробка моделі PLS Calibration, алгоритм класифікації та розгортання під час виконання також включені в пакет. StellarPro має налаштування дозволів користувача, входу та звітності для експорту всіх даних в одному місці. Також включені додаткові програми для дослідження тонких плівок, мутних середовищ, тощо.

StellarNet надає універсальний набір програмного забезпечення Python SDK для взаємодії з лінійкою спектрометрів, що забезпечує інтеграцію та контроль для різноманітних спектроскопічних програм. Цей SDK підтримує збір даних у реальному часі та забезпечує функції для калібрування, спектрального аналізу та візуалізації даних. SDK містить вичерпну документацію та приклади сценаріїв, щоб полегшити швидку розробку та розгортання спеціальних спектроскопічних рішень. Можливість інтеграції з іншими бібліотеками та інструментами Python додатково розширює можливості спектрометрів StellarNet.

Для інтеграції з хмарною платформою використано AWS IoT Core Python SDK, що надає надійний набір інструментів для підключення, керування та взаємодії з пристроями Інтернету Речей через Amazon Web Services. Цей SDK спрощує інтеграцію пристроїв IoT із службами AWS, забезпечуючи збір, обробку та аналіз даних у реальному часі. Завдяки підтримці протоколів MQTT, HTTP та WebSockets, SDK забезпечує надійний і безпечний зв'язок між пристроями та ядром AWS IoT. SDK пропонує функції для публікації повідомлень і підписки, а також безпечного зв'язку за допомогою TLS/SSL. Ця гнучкість дозволяє розробникам створювати масштабовані та ефективні додатки IoT, використовуючи повний спектр сервісів AWS, таких як AWS Lambda, AWS S3 і Amazon DynamoDB.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

Високорівневу архітектуру автоматизованого рішення представлено на рис. 2.

Було розроблено Python скрипт, що інтегрує спектрометр StellarNet із AWS IoT Core, дозволяє отримувати спектрографічні дані в реальному часі та публікувати їх у хмарі. Скрипт призначений для калібрування спектрометра, постійного збору даних спектру та надсилання цих даних до AWS IoT Core для подальшої обробки або аналізу.

Скрипт починається з налаштування необхідних параметрів для підключення до AWS IoT Core. Це включає вказування посилання, ідентифікатора клієнта та шляхів до сертифікатів SSL/TLS, необхідних для безпечного зв'язку. Використовуючи метод ``mqtt_connection_builder.mtls_from_path`` з AWS IoT Core Python SDK, сценарій ініціалізує клієнт MQTT. Потім встановлюється з'єднання з AWS IoT Core, що дозволяє сценарію публікувати дані у вказаному MQTT-топіку. З'єднання MQTT через Transport Layer Security (TLS) забезпечує безпечний і надійний зв'язок між локальним спектрометром і хмарним середовищем AWS.

Дані сценарій ініціалізує спектрометр StellarNet за допомогою бібліотеки ``stellarnet_driver3``. Спектрометр ініціалізується викликом функції ``sn.array_get_spec(0)``, яка отримує об'єкт спектрометра та пов'язані дані про довжину хвилі. Після ініціалізації сценарій переходить до калібрування спектрометра. Функція ``calibrate_spectrometer`` встановлює параметри, такі як час інтеграції, кількість сканувань для усереднення, коефіцієнт згладжування та стан синхронізації зовнішнього тригера. Ці параметри мають вирішальне значення для забезпечення точного та узгодженого збору даних про спектр.

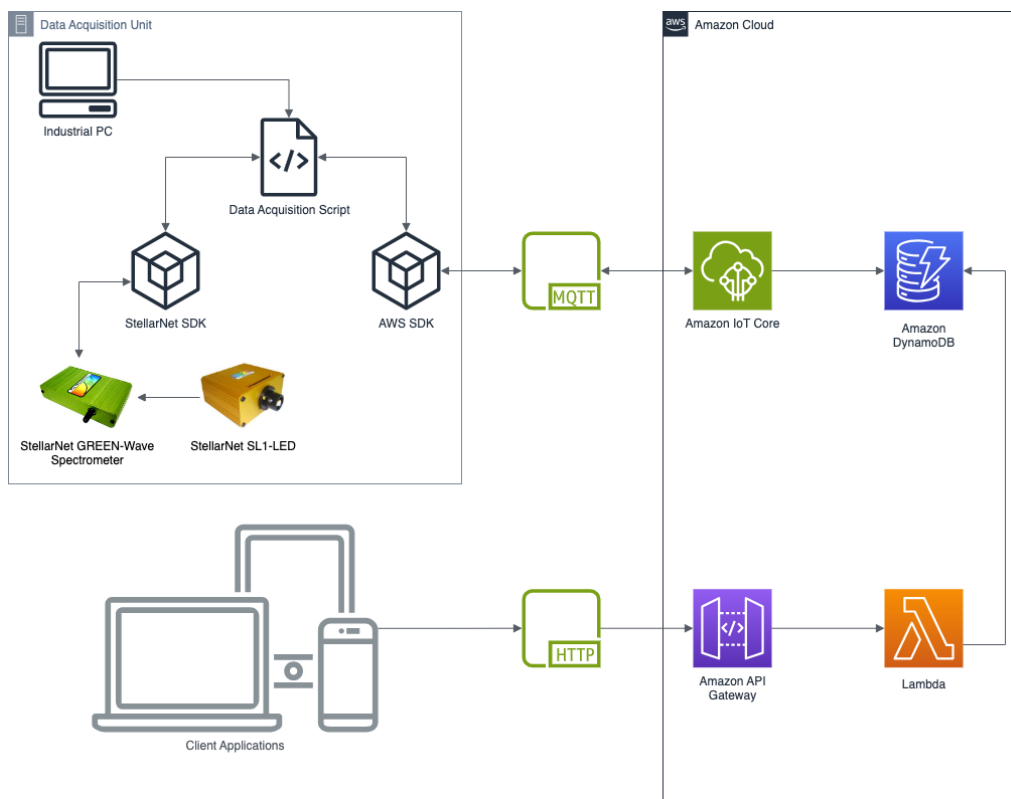


Рис. 2. Архітектура хмарної лабораторії
Fig. 2. Cloud Lab Architecture

Після калібрування сценарій входить у безперервний цикл, де він отримує спектральні дані в реальному часі зі спектрометра. Функція `get_spectrum_data` використовується для отримання спектральних даних - довжини хвилі та відповідних значень інтенсивності. Ця функція викликає метод `sn.array_spectrum(spectrometer, wav)` для отримання поточних даних спектру. Далі дані форматуються у JavaScript Object Notation (JSON) сумісну структуру, включаючи позначку часу (timestamp), що вказує, коли дані були зібрані.

Після отримання спектральних даних вони надсилаються до AWS IoT Core за допомогою функції `publish_to_aws_iot`. Ця функція перетворює дані в JSON і публікує їх у вказаній темі на AWS IoT Core за допомогою методу `mqtt_connection.publish`. Публікуючи дані в режимі реального часу, скрипт дозволяє дистанційно контролювати та аналізувати покази спектрометра. Це налаштування особливо корисно для програм, які вимагають безперервного збору даних і опрацювання в хмарі. Окрім того, скрипт дозволяє підписатись на той же топик AWS IoT Core, що дозволяє перевіряти надіслані дані і відстежувати помилки.

Скрипт починається з калібрування спектрометра, а потім входить у нескінченний цикл, де він отримує та публікує дані спектру кожні 5 секунд (на практиці для конкретних сценаріїв цю затримку можна збільшувати чи зменшувати). Цикл може бути перерваний користувачем (перериванням виконання клавішами Ctrl+C в терміналі), у цей момент сценарій гарантує, що спектрометр належним чином скинуто, а з'єднання MQTT правильно закрито. Це гарантує, що спектрометр залишається в «чистому» стані та те, що немає тривалих підключень до AWS IoT Core, які не використовуються.

Таким чином, скрипт надає комплексне рішення для інтеграції спектрометра StellarNet з AWS IoT Core. Він охоплює ініціалізацію, калібрування, збір даних у реальному часі та публікацію в хмарі. Використовуючи можливості StellarNet SDK та AWS IoT Core SDK, сценарій забезпечує плавний і ефективний моніторинг даних спектрометра, що дозволяє автоматизувати експерименти у лабораторії з підключеними пристроями та спостерігати за результатами у реальному часі. Приклад роботи скрипта показано на рис. 3.

```
Spectrometer connected!
set_ext_trig True
setParam inttime=50 scansavg=1 smooth=0 xtiming=3
setTempComp True
Connection Successful with return code: 0 session present: False
Connected to AWS!
Subscribing to topic 'spectrum'...
Subscribed with QoS.AT_LEAST_ONCE
Received message from topic 'spectrum': b'{"timestamp": 1721052925, "wavelength": [339, 356, 374, 391, 409, 428, 446, 464, 483, 501, 520, 539, 559, 578, 597, 617, 637, 657, 677, 697, 717, 738, 759, 780, 801, 822, 843, 865, 886, 908, 930, 952, 974, 997, 1019, 1042, 1065, 1088, 1111, 1134, 1158], "intensity": [827, -4, -4, 9, -6, 2, -14, 67, -36, -1, 22, -4, -6, -16, 3, -16, -17, 9, -11, -14, -16, -8, 1, -16, -4, -3, -22, 7, 18, 36, 32, 8, 8, -70, -6, -3, 7, -16, -20, -6, 16]}'
Received message from topic 'spectrum': b'{"timestamp": 1721052930, "wavelength": [339, 356, 374, 391, 409, 428, 446, 464, 483, 501, 520, 539, 559, 578, 597, 617, 637, 657, 677, 697, 717, 738, 759, 780, 801, 822, 843, 865, 886, 908, 930, 952, 974, 997, 1019, 1042, 1065, 1088, 1111, 1134, 1158], "intensity": [853, -4, 8, 12, 20, 15, 13, -40, 28, -36, -23, 12, 12, -16, -11, 37, 2, -15, -23, -6, 19, -8, -36, 17, -5, 7, -24, -26, -19, -12, -5, 52, -33, 9, 25, 17, 30, 54, 45, -17, -11]}'
Received message from topic 'spectrum': b'{"timestamp": 1721052936, "wavelength": [339, 356, 374, 391, 409, 428, 446, 464, 483, 501, 520, 539, 559, 578, 597, 617, 637, 657, 677, 697, 717, 738, 759, 780, 801, 822, 843, 865, 886, 908, 930, 952, 974, 997, 1019, 1042, 1065, 1088, 1111, 1134, 1158], "intensity": [825, 40, 20, 4, 42, 2, 77, 607, 11295, 54686, 16874, 3848, 853, 261, 60, 49, 33, 49, 9, 6, 45, -27, -2, 13, -21, 0, 13, 67, -4, 0, 42, 12, 17, -10, 26, 52, 48, 18, -23, -31, 16]}'
Received message from topic 'spectrum': b'{"timestamp": 1721052941, "wavelength": [339, 356, 374, 391, 409, 428, 446, 464, 483, 501, 520, 539, 559, 578, 597, 617, 637, 657, 677, 697, 717, 738, 759, 780, 801, 822, 843, 865, 886, 908, 930, 952, 974, 997, 1019, 1042, 1065, 1088, 1111, 1134, 1158], "intensity": [845, 16, 40, 10, 45, 79, 60, 586, 11214, 54250, 17151, 3947, 901, 212, 92, 12, -23, 4, 25, -23, 49, -5, 64, 12, -15, -39, -3, 28, -34, -54, 18, 11, 41, 5, 61, 45, 33, 22, 11, -8, -10]}'
Disconnecting from AWS...
Connection closed
Disconnected!
Releasing spectrometer...
Released!
```

Рис. 3. Приклад роботи скрипта з вимкненим та увімкненим LED джерелом
Fig. 3. Example of script operation with LED source turned off and on

Розглянемо реалізацію рішення зі сторони хмарного середовища. Для передачі даних використовується протокол MQTT, який є легким протоколом обміну повідомленнями, розробленим для малопотужних пристроїв та ненадійних мереж. Він широко використовується в системах IP для забезпечення ефективного, простого та надійного зв'язку між пристроями. Кожен підключений пристрій може використовувати одну чи декілька тем (topics) для публікування/підписки на повідомлення.

Дані передаються напряму в AWS IoT Core. Це керована хмарна служба від Amazon Web Services, яка дозволяє підключеним пристроям легко та безпечно взаємодіяти з хмарними додатками та іншими пристроями. Вона забезпечує інфраструктуру, необхідну для побудови додатків IP, які збирають, обробляють, аналізують і реагують на дані, згенеровані підключеними пристроями, без необхідності управління базовою інфраструктурою.

У межах сервісу створено Річ (Thing) під назвою "stellarnet". Річ – це представлення та запис фізичного пристрою в хмарному середовищі. Фізичному пристрою потрібен запис про річ, щоб працювати з AWS IoT. Типовий варіант використання пристрою передбачає використання назви речі як ідентифікатора клієнта MQTT за замовчуванням. У роботі використано той самий Client ID "stellarnet" у скрипті, який відправляє дані в хмару.

Рис. 4 показує відображення даних про статус під'єднання та авторизації спектрометра у хмарному середовищі, а також статистику щодо опублікованих повідомлень по темі.

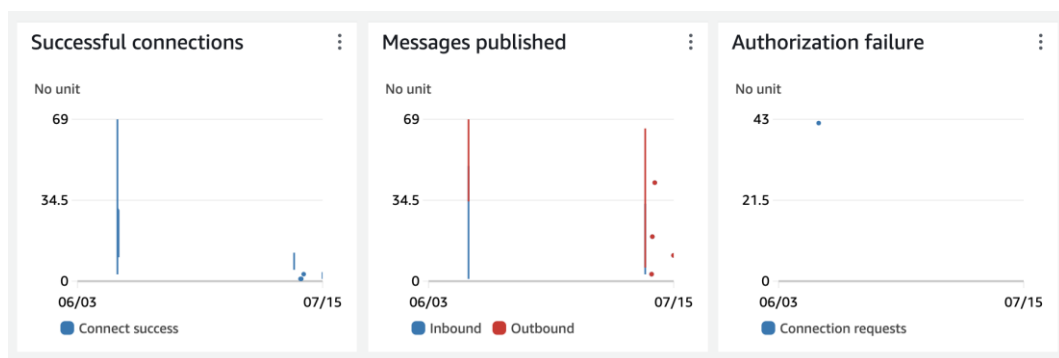


Рис. 4. Приладова панель отриманих та надісланих повідомлень в AWS IoT Core
Fig. 4. Dashboard of received and sent messages in AWS IoT Core

Також в межах AWS IoT Core використовуються правила (Rules). Правило - це компонент AWS IoT Core, який дозволяє автоматично обробляти повідомлення, що надходять від під'єднаних пристроїв. Використовуючи правила, ми можемо фільтрувати, перетворювати і маршрутизувати ці повідомлення до інших сервісів AWS або виконувати інші дії на їхній основі. Зокрема, створено правило під назвою "save_to_db", яке підписується на топик "spectrum", що використовується для публікації даних зі скрипта, описаного вище. Правило вибирає усі дані з повідомлення за допомогою SQL-виразу та передає через конектор до бази даних на AWS DynamoDB.

Для зберігання оброблених даних спектрометра обрано сервіс Amazon DynamoDB завдяки його високій продуктивності та масштабованості. DynamoDB — це повністю керована служба бази даних NoSQL, яка забезпечує гнучке зберігання даних із низькою часовою затримкою. Створено таблицю з атрибутами 'timestamp', 'wavelength' та 'intensity' (кожен наступний пристрій, відповідно, матиме свою таблицю з іншими полями). Завдяки використанню можливостей автоматичного

масштабування DynamoDB рішення може обробляти великий обсяг вхідних даних без ручного втручання, гарантуючи, що рівень зберігання залишається надійним навіть за великих навантажень. Типові записи у базі показано на рис. 5.

timestamp (Number) ▾	intensity ▾	wavelength ▾
1720739443	[{"N": "998"}, {"N": "-10"}, {"N": "-1..."	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777172	[{"N": "876"}, {"N": "31"}, {"N": "10" ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777618	[{"N": "871"}, {"N": "9"}, {"N": "-8"}, ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777212	[{"N": "897"}, {"N": "4"}, {"N": "3"}, {...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720731060	[{"N": "997"}, {"N": "11"}, {"N": "34" ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1721052936	[{"N": "825"}, {"N": "40"}, {"N": "20" ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777608	[{"N": "888"}, {"N": "24"}, {"N": "-13"...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777587	[{"N": "885"}, {"N": "22"}, {"N": "-5" ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720739446	[{"N": "1030"}, {"N": "-2"}, {"N": "-2..."	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720777536	[{"N": "899"}, {"N": "34"}, {"N": "25" ...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720739459	[{"N": "1039"}, {"N": "46"}, {"N": "28..."	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720739458	[{"N": "1040"}, {"N": "-19"}, {"N": "1..."	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...
1720739453	[{"N": "1038"}, {"N": "19"}, {"N": "-3"...	[{"N": "339"}, {"N": "356"}, {"N": "374"}, {"N": "...

Рис. 5. Таблиця з даними спектрометра в AWS DynamoDB
Fig. 5. Spectrometer data table in AWS DynamoDB

Дані, які успішно збережено в базу, мають візуалізуватися. Amazon має кілька сервісів для візуалізації, наприклад QuickSite або SiteWise. AWS QuickSight — це хмарна служба бізнес-аналітики, яка дозволяє не лише візуалізувати дані, а й отримувати аналітичні інсайти. AWS IoT SiteWise — це керована хмарна служба AWS, яка допомагає збирати, організовувати та аналізувати дані з виробничого обладнання. Обидві служби дають можливість створювати дашборди з графіками, але, на жаль, не пристосовані для відображення даних в реальному часі, оскільки спрямовані на роботу з аналітикою та агрегацію даних за певний період. Тому було вирішено розробити окремий додаток для візуалізації даних з DynamoDB.

Для доступу до даних використано сервіс AWS API Gateway. Це керована служба, яка дозволяє розробникам створювати, публікувати, підтримувати, моніторити та захищати API (Application Programming Interfaces) будь-якого масштабу. Сервіс також дозволяє створювати та керувати RESTful або WebSocket API для застосунків, що взаємодіють із бекенд-сервісами. Було створено REST API під назвою "spectrum" з ендпоинтом GET /latest. Згенеровано URL адресу, за допомогою якої можна звертатись до цього ендпоинта. Ендпоинт GET /latest, у свою чергу, прив'язаний до інтеграції з AWS Lambda function "spectrum".

AWS Lambda - це serverless обчислювальна служба, яка дозволяє запускати код без необхідності створення сервера на віртуальній машині. AWS Lambda виконує код тільки тоді, коли це потрібно, і масштабується автоматично від декількох запитів на день до тисяч на секунду. Було створено функцію "spectrum", яка перевіряє вхідний routeKey і обробляє GET /latest запит з API Gateway. Функція отримує дані з бази, вибирає останній елемент і перетворює дані з формату DynamoDB у JSON, який відправляється у відповідь на запит (рис. 6).

```

import json
import boto3
from decimal import Decimal

client = boto3.client('dynamodb')
dynamodb = boto3.resource("dynamodb")
table = dynamodb.Table('spectrum')
tableName = 'spectrum'

def lambda_handler(event, context):
    print(event)
    body = {}
    statusCode = 200
    headers = { "Content-Type": "application/json" }
    try:
        if event['routeKey'] == "GET /latest":
            latest = table.scan()["Items"][-1]
            intensity = []
            for i in latest['intensity']:
                intensity.append(int(i))
            wavelength = []
            for i in latest['wavelength']:
                wavelength.append(int(i))
            body = {
                'timestamp': int(latest['timestamp']),
                'intensity': intensity,
                'wavelength': wavelength
            }
        except KeyError:
            statusCode = 400
            body = 'Unsupported route: ' + event['routeKey']
    body = json.dumps(body)

```

Рис. 6. AWS Lambda функція для зв'язування API Gateway та DynamoDB
 Fig. 6. AWS Lambda function for linking API Gateway and DynamoDB

Для реалізації додатку візуалізації даних було використано Flutter. Це відкритий фреймворк для створення кросплатформних мобільних, веб та десктопних додатків, розроблений компанією Google. Він дозволяє розробникам створювати додатки для iOS, Android, веб та настільних платформ, використовуючи єдину базу коду. Flutter використовує власний графічний рушій Skia для візуалізації, що забезпечує високу продуктивність та швидкість роботи додатків. А завдяки функції "Hot Reload" розробники можуть миттєво бачити зміни у своєму коді без необхідності перезавантажувати додаток, що значно прискорює процес розробки.

У рамках роботи розроблено додаток "spectroscopy_app" (рис. 7), що отримує дані з AWS API Gateway, описаного вище, і візуалізує їх за допомогою бібліотеки FL Chart (https://pub.dev/packages/fl_chart). FL Chart — це бібліотека для побудови графіків та діаграм для Flutter із широкими можливостями налаштування, яка підтримує лінійну діаграму, гістограму, секторну, точкову та радарну діаграми. Застосунок надсилає GET /latest запит на Web API кожні 5 секунд, після чого оновлює отримані дані на графіку.

В подальшому планується також розробка графіка залежності домінантної частоти від часу. Сюди ж будуть додаватись графіки та приладові панелі (dashboards) з інших підключених пристроїв IP.

ВИСНОВКИ

Автоматизація експериментів з використанням описаного хмарного рішення не тільки підвищує ефективність, але й забезпечує вищу точність і відтворюваність у наукових дослідженнях. Зводячи до мінімуму ручне втручання та використовуючи обробку даних у реальному часі та розширену аналітику, дослідники можуть зосередитися на розробці експериментів та інтерпретації результатів, зрештою прискорюючи темп відкриттів та інновацій.

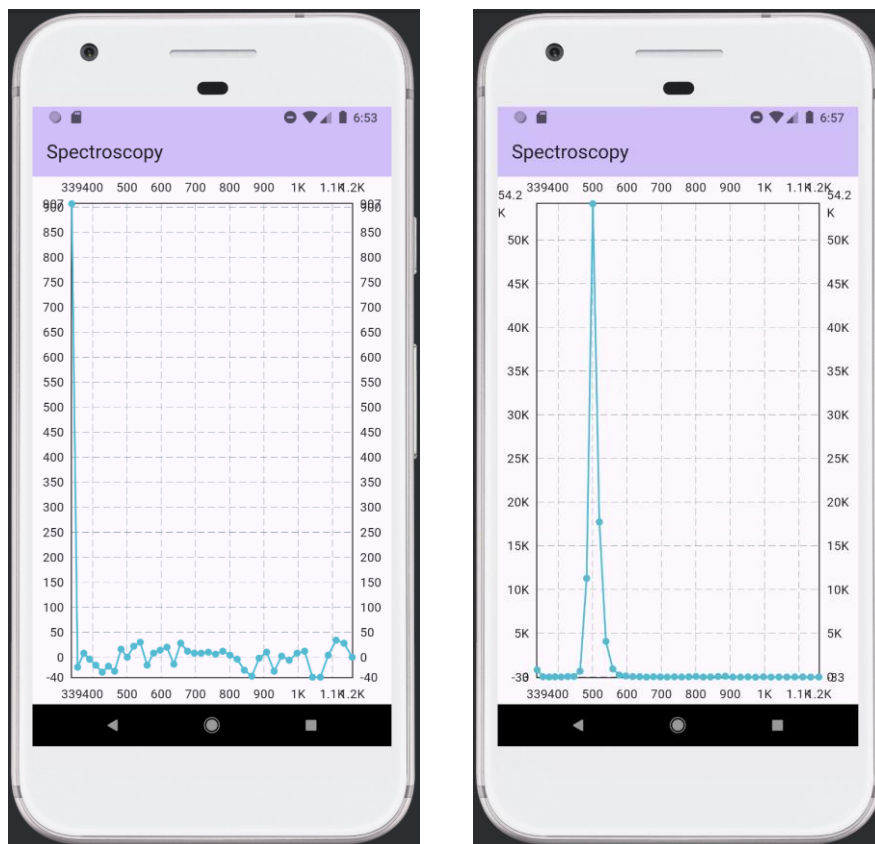


Рис. 7. Візуалізація спектра в мобільному додатку з вимкнутим та увімкнутим LED-джерелом.
Fig. 7. Spectrum visualization in a mobile application with the LED source turned off and on.

Запропоноване інтегроване рішення ефективно використовує потужність сервісів AWS для підвищення функціональності та зручності використання спектрометрів StellarNet. З використанням AWS IoT Core, AWS Lambda, Amazon DynamoDB і Amazon API Gateway створено надійну, масштабовану та ефективну систему для збору, обробки та зберігання даних. Безперебійна взаємодія між цими службами гарантує, що дані спектрометра не тільки збираються та безпечно зберігаються, але й легко доступні для аналізу та прийняття рішень. Додатково розроблений користувацький крос-платформний додаток на Flutter дозволяє переглядати результати експерименту в реальному часі та агрегувати їх за потреби.

Загалом, рішення демонструє потенціал хмарних сервісів у вдосконаленні традиційного лабораторного обладнання. Інтегруючи сучасні хмарні технології з науковими інструментами, дослідники та вчені можуть досягти більшої ефективності, масштабованості та розуміння даних. Цей підхід не тільки оптимізує використання наявних ресурсів, але й прокладає шлях до більш досконалого комп'ютерного опрацювання даних і робочих процесів аналізу в майбутньому.

ВНЕСОК АВТОРІВ

Концептуалізація, [А.К., І.К.]; методологія, [А.К., І.К.]; валідація, [І.К.]; формальний аналіз, [А.К., І.К.]; дослідження, [А.К., І.К.]; написання – підготовка оригінального тексту, [А.К.]; написання – рецензування та редагування, [І.К.]; візуалізація, [А.К.]; нагляд, [І.К.]; адміністративне управління проектом, [І.К.].

Усі автори ознайомилися з остаточною версією рукопису та погодилися з його публікацією.

ДЖЕРЕЛА

- [1] Azad, A. Design and Development of Remote Laboratories with Internet of Things Setting. *Advances in Internet of Things*, 11, 95-112 (2021). DOI: 10.4236/ait.2021.113007.
 - [2] Hairuddin, M., Ashar, N., Abidin, A., Tahir, N. Cost-Effective Interfaces with Arduino-LabVIEW for an IOT-Based Remote Monitoring Application. *LabVIEW - A Flexible Environment for Modeling and Daily Laboratory Use* (2021). doi: 10.5772/intechopen.97784
 - [3] D.F. Parks, et al. IoT cloud laboratory: Internet of Things architecture for cellular biology. *Internet of Things*, 20 (2022), Article 100618. DOI: 10.1016/j.iot.2022.100618
 - [4] M. Domínguez, R. González-Herbón, J.R. Rodríguez-Ossorio, J.J. Fuertes, M.A. Prada and A Morán. Development of a remote industrial laboratory for automatic control based on Node-RED. *IFAC-PapersOnLine* (2020), vol. 53, no. 2, pp. 17210-17215. DOI: 10.1016/j.ifacol.2020.12.1741
-

COMPUTERIZED OPTICAL EXPERIMENTS WITH PORTABLE SPECTROMETERS AND AMAZON WEB SERVICES CLOUD INTEGRATION

Andriy Krupych[✉], **Ivan Karbovnyk**[✉]
Ivan Franko National University of Lviv
Drahomanova Str., 50, 79005 Lviv, Ukraine

ABSTRACT

Background. This paper presents a system for automating optical experiments using the StellarNet spectrometer and AWS cloud services. The relevance of this research is driven by the need for remote access to laboratory equipment, which became particularly critical during the COVID-19 pandemic. The proposed solution enables real-time optical experiments, automates spectral data collection and processing, and ensures secure cloud storage for further analysis. By leveraging cloud technologies, the system enhances research efficiency and accessibility, reducing dependence on physical presence in the laboratory.

Materials and Methods. The study reviews modern approaches to laboratory automation, focusing on both hardware and software solutions. The integration of embedded processors like Raspberry Pi, development environments such as LabVIEW, and IoT-based architectures is considered. A key aspect is the implementation of remote access, allowing seamless operation and real-time monitoring. The system hardware includes the StellarNet VIS-50 spectrometer, the SL1-LED light source, and the Synco PC Box mini-PC. Software tools, such as the StellarNet Python SDK and AWS IoT Core, are used to enable automated data processing and cloud communication.

Results and Discussion. A Python script was developed for integrating the spectrometer with AWS IoT Core, enabling seamless data transmission. The paper describes the setup, calibration, and real-time spectral data collection process. Using the MQTT protocol, experimental data is securely transmitted to the cloud, allowing remote access and analysis. The implementation details, including secure connection protocols,

message publication, and cloud integration, are discussed to highlight system reliability and efficiency.

Conclusions. The proposed system enhances laboratory experiment automation, improving research flexibility and efficiency in STEM disciplines. It reduces equipment maintenance costs and increases accessibility to experimental data. The scalable solution can be adapted to various scientific applications requiring remote equipment access and continuous data collection.

Keywords: Automation, Data Processing, Internet of Things, Spectroscopy, AWS, Cloud Computing

UDC 621.382, 537.312

PHOTOSENSITIVE STRUCTURES BASED ON THE SILVER-DOPED POROUS SILICON AND REDUCED GRAPHENE OXIDE

Igor Olenych , Lyubomyr Monastyrskii , Petro Parandiy ,
Roman Serkiz , Ivan Karbovnyk 
Ivan Franko National University of Lviv
50 Dragomanov St., Lviv 79005, Ukraine

Olenych, I. et al. (2025). Photosensitive structures based on the silver-doped porous silicon and reduced graphene oxide. Electronics and Information Technologies, 29, 135-144. <https://doi.org/10.30970/eli.29.12>

ABSTRACT

Background. The synergistic combination of useful properties of the different nature nanostructures in composites is one of the main methods of improving the functionality of nanomaterials. In particular, due to the unique properties of graphene as well as the large absorbing surface area and the low reflectance of porous silicon (PS), hybrid structures based on them are promising for photoelectric applications. Therefore, the possibility of increasing the photosensitivity of such structures by electrochemical introduction of silver nanoparticles into the porous layer was studied.

Materials and Methods. Nanostructured layers of the PS obtained by electrochemical etching of a silicon wafer and a film-forming suspension of reduced graphene oxide (RGO), which was deposited on the porous layer surface, were used to create photosensitive structures. The resulting structures with silver incorporated into the PS were explored using scanning electron microscopy. The electrical and photoelectric properties of the hybrid structures were studied based on the analysis of I - V characteristics and impedance frequency dependencies measured in the dark and under irradiation with white light.

Results and Discussion. It has been established that silver is deposited mainly in the surface layer of the PS and its amount depends on the duration of electrochemical introduction. Doping the porous layer with silver increases the electrical conductivity and capacitance of the PS–RGO structures. Irradiation of the surface of experimental structures with white light causes an increase in the forward and reverse current through the structures but has almost no effect on the lateral I - V characteristics due to shunting the RGO film by silver nanoparticles. Based on the analysis of impedance spectra in the 10^2 – 10^5 Hz frequency range, the effect of irradiation on the resistive-capacitive properties of hybrid structures was studied.

Conclusion. It was established that the incorporated silver forms additional current paths through the PS. The RGO film resistance is partially shunted by the silver nanoparticles as well. As a result, the obtained structures demonstrate an increase in conductivity both when the charge is transferred through the structure and along the surface. In addition, silver nanoparticles contribute to the accumulation of photogenerated charge carriers and their extraction from the PS.

Keywords: Porous silicon, reduced graphene oxide, silver nanoparticles, photosensitivity, impedance.



© 2025 Igor Olenych et al. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and information technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

An important area of R&D in nanoelectronics is improving the functionality of nanomaterials. The combination of useful properties of the different nature nanostructures in composite materials is one of the main ways to achieve this goal. Due to the synergistic effect, nanocomposites often demonstrate improved or completely new properties and therefore have a high potential for use in devices of various appointments. Special attention of researchers is focused on hybrid materials based on graphene, which have the prospect of application in solar cells, energy storage devices, sensors, medicine and the food industry [1-4].

The outstanding properties of graphene are due to the conical shape of the electronic spectrum of the hexagonal structure of a monolayer of sp^2 -bonded carbon atoms [5, 6]. As a result, 2D graphene is characterized by high charge carrier mobility, high thermal conductivity, and the ability to withstand significant current density, which is a significant advantage for creating high-speed and powerful electronic devices [7-9]. The low electrical noise of the carbon monolayer provides a high sensitivity of graphene-based nanocomposites to the local electric field of adsorbed molecules [10]. Besides, graphene can be used as a transparent electrode in photoelectric devices due to its light absorption of only 2.3 % [11]. However, the high transparency of graphene and the high recombination rate of photoinduced charge carriers cause the weak photoresponse of this nanomaterial. One of the main strategies for improving the photosensitivity of carbon nanostructures is hybridization: the combination of graphene and light-sensitive semiconductors [12, 13]. In particular, hybrid structures based on reduced graphene oxide (RGO) film and porous silicon (PS) are promising for photovoltaic applications [14] because the PS layers are characterized by a large absorbing surface area and excellent anti-reflective properties [15, 16]. However, the high resistance of the PS imposes some limitations on the application of the material in photoelectronics, which requires finding ways to increase the efficiency of charge transfer through the porous layer.

This study aimed to explore the influence of electrochemical doping of the porous layer with silver on the electrical and photoelectric properties of the hybrid structures based on the PS and RGO film.

MATERIALS AND METHODS

The PS layer for photosensitive hybrid structures was obtained by electrochemical etching of a 400 μm thick silicon wafer of electronic conductivity type with a 4.5 $\text{Ohm}\cdot\text{cm}$ resistivity. An ethanol solution of hydrofluoric acid with the component ratio of $\text{HF}:\text{C}_2\text{O}_5\text{OH} = 1:1$ was used as the electrolyte. The anodic current density and etching time were 30 mA/cm^2 and 10 min, respectively. The working surface of the silicon wafer was irradiated with a 500 W lamp to increase the etching efficiency. A thin gold film that served as an electrical contact was previously deposited on the silicon substrate back surface by thermal technique and annealed at a 600 $^\circ\text{C}$ temperature for 20 min.

The doping of the PS layer with silver was carried out from an aqueous solution of silver nitrate (AgNO_3) using the electrochemical method. The value of the direct current was 1 mA for 5, 10 and 20 s to obtain different experimental samples. The resulting structures are labeled as sample 1, sample 2, and sample 3, respectively. In the case of passing the electric current through the electrolyte, metal ions are transferred to the cathode and deposited not only on the surface but also in the pores of the PS. Silver ions, which have a more positive electrochemical potential than silicon, after depositing on the surface of the walls between the pores are neutralized by withdrawing electrons from surface silicon atoms and become the nuclei for the growth of silver nanocrystals. In addition, redox processes can contribute to the formation of an oxide film on the PS surface. The incorporation of silver into the porous layer was controlled using scanning electron

microscopy (SEM) in the elastically reflected electron mode and energy-dispersive X-ray microanalysis (EDXMA).

After washing the formed structures with distilled water, a film-forming RGO suspension obtained by chemical reduction of graphene oxide as described in [17] was deposited on the PS layer surface. Further air-drying of the experimental samples at room temperature ensured the formation of the RGO film with a thickness of up to several tens of nanometers [14, 17]. Two silver contacts were thermally deposited on the RGO film surface at a 1 mm distance from each other.

The electrical and photoelectric properties of the obtained structures based on the Ag-doped PS and RGO film were studied in DC and AC modes using the Siglent SDM 3045 multimeter and the Hantek 1833C RLC-measuring device. Photoelectric phenomena were investigated under the condition of irradiation of the experimental structures from the RGO film side with a white light LED FYLP-1W-UWB-A with a 1 W power and a luminous flux of 76 lumens. All the measurements were carried out at room temperature

RESULTS AND DISCUSSION

The morphology of hybrid structures based on the Ag-doped PS and RGO film was studied using SEM in the secondary electron mode. As a result of electrochemical etching, the formation of long narrow pores directed into the depth of the silicon wafer perpendicular to the surface was observed (Fig. 1). The pore diameter ranged from tens to hundreds of nanometers, which facilitated the effective penetration of the electrolyte into the pores during the silver deposition process. The walls between the pores were also characterized by a nanometer thickness.

In the elastically reflected electron mode, the image areas with higher brightness are identified as formed silver clusters. Analysis of SEM images of the cross-section of the experimental structures made it possible to establish that a significant amount of

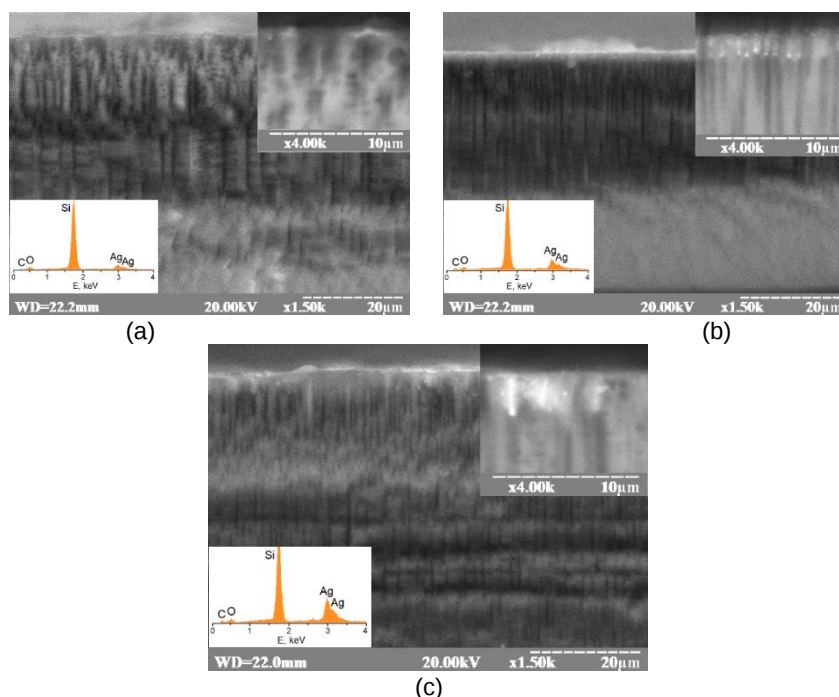


Fig. 1. SEM images of cross-sections of structures based on the Ag-doped PS and RGO in the secondary electron mode: sample 1 (a), sample 2 (b) and sample 3 (c). The insets show SEM images in the elastically reflected electron mode and EDXMA spectra of the corresponding structures.

electrodeposited silver was located in the near-surface layer of the PS. The surface of the walls between the pores in the upper part of the PS layer is well decorated with metal particles. An increase in the amount of incorporated silver is observed with an increase in the time of its electrochemical introduction into the porous layer. An intense maximum with energy of 1.74 keV associated with silicon atoms was detected in the EDXMA spectra. In addition, peaks at 2.98 and 3.15 keV that correspond to silver atoms were identified. The intensity of these peaks was greater for samples with higher amounts of metal. Small bands with energies of 0.3 and 0.5 keV are related to carbon and oxygen atoms, respectively, and can be associated with RGO and the native oxide film on the surface of the PS nanostructures.

Doping the porous layer with silver significantly affects the charge transfer processes in hybrid structures based on the PS and RGO both along the surface of the structures and in the direction perpendicular to the surface. In the first case, the main influence on charge transfer is the electrical properties of the RGO film because the resistance of the PS layer significantly exceeds the resistance of graphene.

As can be seen in Fig. 2, the lateral $I_{||}$ - $V_{||}$ characteristics are linear at small bias voltages. It is worth noting that the $I_{||}$ - $V_{||}$ characteristics are determined not only by charge carriers injected into the RGO film from electrical contacts but also from the PS layer and silver nanoparticles. An increase in the amount of incorporated silver causes a decrease in lateral resistance several times, which may be due to the formation of percolation clusters of metal nanoparticles that partially shunt the RGO film. A further increase in the duration of the silver electrochemical introduction probably causes the formation of additional current paths along the structure surface. As a result, a three-order decrease in the resistance between the contacts on the RGO film surface is observed (see the inset of Fig. 2).

The $I_{\perp}$ - $V_{\perp}$ characteristics of the Ag-doped PS-RGO structures have a nonlinear nature in the case when the current passes perpendicular to the surface (Fig. 3). An increase in the amount of silver incorporated into the porous layer causes increasing the forward current, which corresponds to a positive potential on the RGO film. The duration of doping the PS with silver has almost no effect on the magnitude of the reverse current in the dark. The observed shape of the $I_{\perp}$ - $V_{\perp}$ curves indicates the presence of several electrical barriers in the studied structures. Considering the Ohmic contact of graphene with metals, the dominant barrier is likely to be at the interface of the PS with silver nanoparticles that connect with the RGO film.

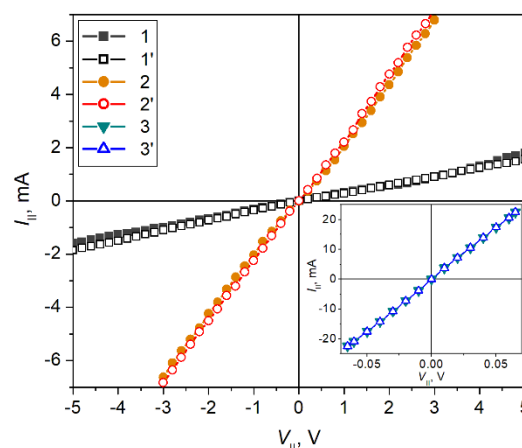


Fig 2. Lateral $I_{||}$ - $V_{||}$ characteristics of the Ag-doped PS-RGO structures obtained in the dark (curves 1, 2 and 3) and under irradiation with white light (curves 1', 2' and 3'). Curves (1, 1'), (2, 2') and (3, 3') related to samples 1, 2 and 3, respectively.

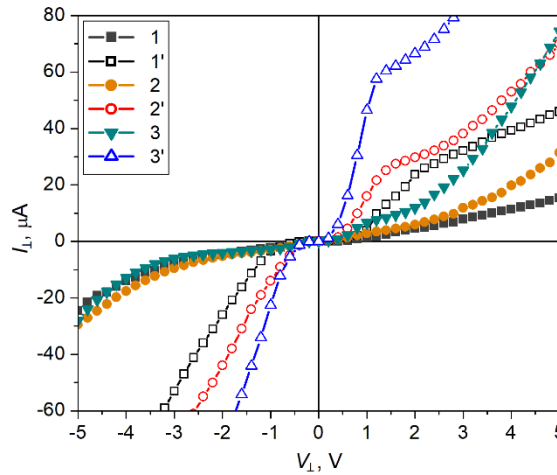


Fig 3. The $I_{\perp}$ - $V_{\perp}$ characteristics of the Ag-doped PS-RGO structures obtained in the dark (curves 1, 2 and 3) and under irradiation with white light (curves 1', 2' and 3'). Curves (1, 1'), (2, 2') and (3, 3') related to samples 1, 2 and 3, respectively.

An increase in both the forward and reverse current $I_{\perp}$ is observed due to photogenerated charge carriers when the Ag-doped PS-RGO structures are irradiated with white light. Experimental samples with a higher amount of incorporated silver demonstrate higher photocurrent values. The incorporation of silver provides an increase in the efficiency of photoelectric structures based on the PS and RGO due to the high electron work function and a larger Schottky barrier area. In addition, silver nanoparticles can better collect light-induced charge carriers and extract them from the porous layer. In the forward current mode, the maximum influence of photocarriers on the $I_{\perp}$ - $V_{\perp}$ curves is observed at the small bias voltage (up to 1.2–2.0 V for various samples), after which the shape of the current-voltage dependencies becomes similar to ones in the dark. Identified features of the $I_{\perp}$ - $V_{\perp}$ characteristics may be an additional argument in favor of the assumption of the participation of incorporated silver in the processes of accumulation of photogenerated charge carriers.

It is worth noting that the lateral III-VII characteristics of hybrid structures almost do not change when irradiated with the white light LED (see Fig. 2). This fact may be due to the high transparency of the RGO film and the low-quality dielectric properties of the silver-doped PS layer if the studied structures are used as graphene field-effect transistors. Moreover, silver nanoparticles not only shunt the RGO film but can also block the local electric field of charge carriers that are photogenerated in the PS layer and the silicon substrate.

The frequency dependences of the impedance between the RGO film and the silicon substrate were measured to study the capacitive-resistive properties of the Ag-doped PS-RGO structures. A decrease in capacitance and internal resistance with increasing frequency was observed in the 10^2 – 10^5 Hz range (Fig. 4).

The different dispersion of the electrical parameters of the experimental samples in various frequency ranges along with the identified features of the $I_{\perp}$ - $V_{\perp}$ characteristics indicate complex processes of charge transfer through the Ag-doped PS-RGO structure perpendicular to the surface. Usually, low-frequency and high-frequency parts of impedance dependence are associated with the transfer of charge through the interfaces and in the bulk of nanostructures, respectively. Nyquist plots were used to analyze frequency dependencies of the impedance of the experimental structures (Fig. 5).

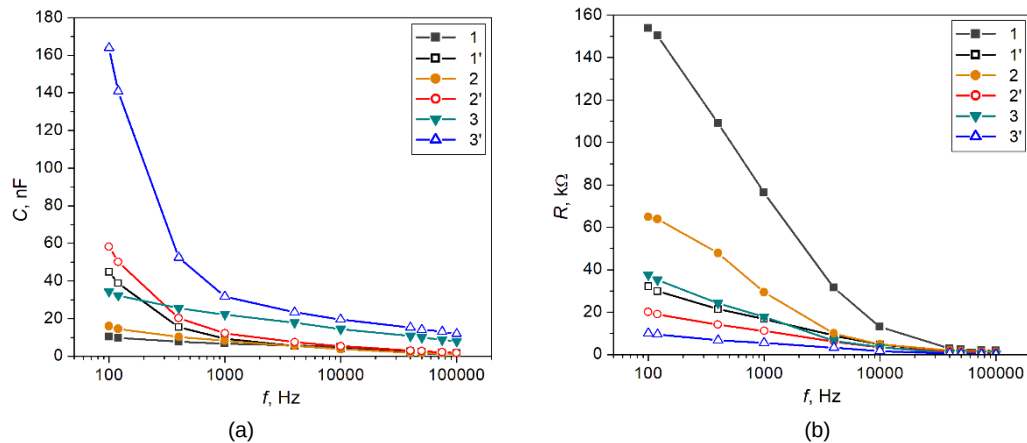


Fig. 4. Frequency dependencies of capacitance (a) and internal resistance (b) of the Ag-doped PS-RGO structures obtained in the dark (curves 1, 2 and 3) and under irradiation with white light (curves 1', 2' and 3'). Curves (1, 1'), (2, 2') and (3, 3') related to samples 1, 2 and 3, respectively.

A parallel-connected constant phase element (CPE) and resistor with the resistance R were applied in the equivalent circuit model considering the capacitive-resistive properties of fractal and composite materials [14, 18]. The impedance of the R-CPE model is defined by the equation

$$Z(\omega) = \frac{R}{1 + (j\omega)^n RQ}. \quad (1)$$

Here Q is the CPE and n characterizes the heterogeneity of the electrical properties of the Ag-doped PS-RGO structures. The parameters of approximation of the impedance spectra in the 400 Hz – 100 kHz frequency range are shown in Table 1.

Based on the analysis of the approximation results, it can be concluded that the capacitive-resistive properties of the PS-RGO hybrid structures depend on the amount of incorporated silver in the porous layer. Increasing the duration of the doping process

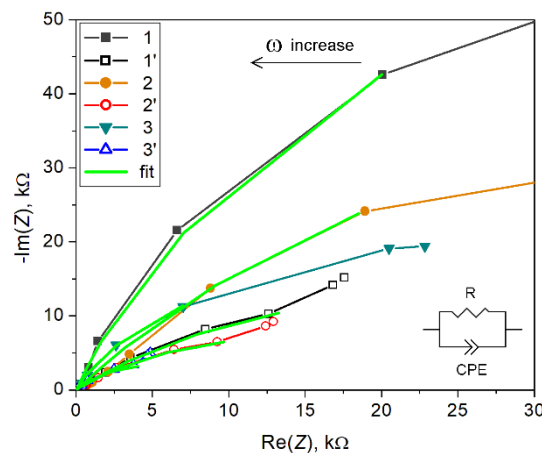


Fig 5. Nyquist plots and an equivalent circuit diagram of the Ag-doped PS-RGO structures obtained in the dark (curves 1, 2 and 3) and under irradiation with white light (curves 1', 2' and 3'). Curves (1, 1'), (2, 2') and (3, 3') related to samples 1, 2 and 3, respectively.

Table 1. The parameters of approximation of the impedance spectra of the Ag-doped PS–RGO structures

Ag-doped PS–RGO structures	Parameters of approximation					
	In the dark			Under irradiation		
	R , k Ω	$Q \times 10^{-7}$	n	R , k Ω	$Q \times 10^{-7}$	n
Sample 1	206.64	0.23	0.86	51.59	4.34	0.59
Sample 2	139.75	1.18	0.70	28.19	5.67	0.59
Sample 3	44.34	0.99	0.83	13.08	6.70	0.68

causes an increase in the conductivity and capacitance of the studied structures due to the formation of additional current paths through the PS layer and a larger area of the Schottky barrier.

Besides, the Ag-doped PS–RGO structures demonstrate an increase in capacitance and a decrease in internal resistance under white light illumination from the LED FYLP–1W–UWB–A.

CONCLUSION

In the work, photosensitive structures were created by depositing a film-forming suspension of RGO onto the PS surface and further air-drying at room temperature. An incorporation of silver nanoparticles into the porous layer was proposed to increase the efficiency of photovoltaic structures. Using SEM methods, it was demonstrated that silver is deposited mainly in the surface layer of the PS and its amount depends on the duration of electrochemical introduction.

Based on the measurement of the I–V characteristics, it was found that the incorporated silver forms additional current paths. As a result, the obtained structures demonstrate an increase in electrical conductivity both when the charge is transferred through the structure and along the surface. An increase in forward and reverse current through the Ag-doped PS–RGO structure was detected under the influence of white light irradiation. However, illumination of the hybrid structure surface has almost no effect on the lateral I–V characteristics due to the RGO film shunting and blocking of the local electrical field of photogenerated charges by silver nanoparticles. An increase in capacitance and a decrease in internal resistance between the RGO film and the silicon substrate both due to increasing the duration of electrochemical incorporation of silver and white light exposure have been established based on the analysis of the impedance frequency dependencies in the 10^2 – 10^5 Hz range. The obtained results can be used to create photodetectors based on hybrid nanomaterials.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, [I.O., L.M., P.P.]; methodology, [I.O., L.M., P.P.]; investigation, [I.O., P.P., R.S., I.K.]; writing – original draft preparation, [I.O.]; writing – review and editing, [I.O., L.M., P.P.]; visualization, [I.O., R.S., I.K.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Mohapatra, P., Behera, S., Sahoo, S., Mishra, A., Kesh, A., Shubhadarshinee, L., Jali, B. R., Mohapatra, P., & Barick, A. K. (2024). Synergistic effect of silver nanoparticles decorated graphene quantum dots nanohybrids reinforced polyaniline ternary nanocomposites on optical, thermal, and dielectric properties. *Composite Interfaces*, 32(1), 1–26. <https://doi.org/10.1080/09276440.2024.2374102>
- [2] Chen, W., Song, K., Mi, L., Feng, X., Zhang, J., Cui, S., & Liu, C. (2017). Synergistic effect induced ultrafine SnO₂/graphene nanocomposite as an advanced lithium/sodium-ion batteries anode. *J. Mater. Chem. A*, 5, 10027–10038. <https://doi.org/10.1039/C7TA01634D>
- [3] Toto, E., Laurenzi, S., & Santonicola, M. G. (2022). Recent Trends in Graphene/Polymer Nanocomposites for Sensing Devices: Synthesis and Applications in Environmental and Human Health Monitoring. *Polymers*, 14(5), 1030. <https://doi.org/10.3390/polym14051030>
- [4] Liew, W. C., Muhamad, I. I., Chew, J. W., & Karim, K. J. A. (2023). Synergistic effect of graphene oxide/zinc oxide nanocomposites on polylactic acid-based active packaging film: Properties, release kinetics and antimicrobial efficiency. *International Journal of Biological Macromolecules*, 253(6), 127288. <https://doi.org/10.1016/j.ijbiomac.2023.127288>
- [5] Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Katsnelson, M. I., Grigorieva, I. V., Dubonos, S. V., & Firsov, A. A. (2005). Two-dimensional gas of massless Dirac fermions in graphene. *Nature*, 438, 197–200. <https://doi.org/10.1038/nature04233>
- [6] Zhan, D., Yan, J., Lai, L., Ni, Z., Liu, L., & Shen, Z. (2012). Engineering the electronic structure of graphene. *Adv. Mater.*, 24(30), 4055–4069. <https://doi.org/10.1002/adma.201200011>
- [7] Moon, J. S. (2012). Graphene Field-effect transistor for radio-frequency applications: review. *Carbon Letters*, 13, 17–22. <https://doi.org/10.5714/CL.2012.13.1.017>
- [8] Xia, F., Perebeinos, V., Lin, Y. M., Wu, Y., & Avouris, P. (2011). The origins and limits of metal-graphene junction resistance. *Nature Nanotech.*, 6, 179–184. <https://doi.org/10.1038/nnano.2011.6>
- [9] Shahil, K. M. F., & Balandin, A. A. (2012). Thermal properties of graphene and multilayer graphene: Applications in thermal interface materials. *Solid State Commun.*, 152, 1331–1340. <https://doi.org/10.1016/j.ssc.2012.04.034>
- [10] Olenych, I. B., Aksimentyeva, O. I., Horbenko, Y. Y., & Tsizh, B. R. (2022). Electrical and sensory properties of silicon–graphene nanosystems. *Appl. Nanosci.*, 12, 579–584. <https://doi.org/10.1007/s13204-021-01698-7>
- [11] Avouris, P. (2010). Graphene: Electronic and Photonic Properties and Devices. *Nano Letters*, 10, 4285–4294. <https://doi.org/10.1021/nl102824h>
- [12] Roy, K., Padmanabhan, M., Goswami, S., Sai, T. P., Ramalingam, G., Raghavan, S., & Ghosh, A. (2013). Graphene–MoS₂ hybrid structures for multifunctional photoresponsive memory devices. *Nature Nanotech.*, 8, 826–830. <https://doi.org/10.1038/nnano.2013.206>
- [13] Palanisamy, T., Mitra, S., Batra, N., Smajic, J., Emwas, A. H., Roqan, I., & Costa, P. M. F. J. (2022). Carbon Nitride Thin Film-Sensitized Graphene Field-Effect Transistor: A Visible-Blind Ultraviolet Photodetector. *Advanced Materials Interfaces*, 9, 2200313. <https://doi.org/10.1002/admi.202200313>
- [14] Olenych, I. B., Aksimentyeva, O. I., Monastyrskii, L. S., Horbenko, Y. Y., & Partyka, M. V. (2017). Electrical and photoelectrical properties of reduced graphene oxide – porous silicon nanostructures. *Nanoscale Res. Lett.*, 12, 272. <https://doi.org/10.1186/s11671-017-2043-7>

- [15] Bisi, O., Ossicini, S., & Pavesi, L. (2000). Porous silicon: a quantum sponge structure for silicon based optoelectronics. *Surf. Sci. Rep.*, 38, 1. [https://doi.org/10.1016/S0167-5729\(99\)00012-6](https://doi.org/10.1016/S0167-5729(99)00012-6)
- [16] Föll, H., Christophersen, M., Carstensen, J., & Hasse, G. (2002). Formation and application of porous silicon. *Materials Science and Engineering: R: Reports*, 39(4), 93–141. [https://doi.org/10.1016/S0927-796X\(02\)00090-6](https://doi.org/10.1016/S0927-796X(02)00090-6)
- [17] Olenych, I. B., Monastyrskii, L. S., Aksimentyeva, O. I., Orovcik, L., & Salamakha, M. Y. (2018). Charge transport in porous silicon/graphene-based nanostructures. *Mol. Cryst. Liq. Cryst.*, 673, 32–38. <https://doi.org/10.1080/15421406.2019.1578491>
- [18] Karbovnyk, I., Klym, H., Chalyy, D., Zhydenko, I., & Lukashevych, D. (2022). *Appl. Nanosci.*, 12, 1263 <https://doi.org/10.1007/s13204-021-01810-x>
-

ФОТОЧУТЛИВІ СТРУКТУРИ НА ОСНОВІ ЛЕГОВАНОГО СРІБЛОМ ПОРУВАТОГО КРЕМНІЮ ТА ВІДНОВЛЕНОГО ОКСИДУ ГРАФЕНУ

**Ігор Оленич, Любомир Монастирський, Петро Парандій,
Роман Серкіз, Іван Карбовник**

*Львівський національний університет імені Івана Франка,
вул. Драгоманова 50, 79005 м. Львів, Україна*

АНОТАЦІЯ

Вступ. Синергетичне поєднання корисних властивостей наноструктур різної природи у композитах є одним із основних методів підвищення функціональності наноматеріалів. Зокрема, завдяки унікальним властивостям графену, великій площі поглинаючої поверхні та низькому коефіцієнту відбивання поруватого кремнію гібридні структури на їх основі є перспективними для фотоелектричних застосувань. Тому у роботі вивчена можливість підвищення фоточутливості таких структур шляхом електрохімічного впровадження у поруватий шар наночастинок срібла.

Матеріали та методи. Для створення фоточутливих структур використано наноструктуровані шари поруватого кремнію, одержаного електрохімічним травленням кремнієвої пластини, та плівкоутворювальну суспензію відновленого оксиду графену, яку осаджували на поверхню поруватого шару. Отримані структури з інкорпорованим у поруватий шар сріблом досліджено методами скануючої електронної мікроскопії. Електричні та фотоелектричні властивості гібридних структур досліджено на основі аналізу вольт-амперних характеристик і частотних залежностей імпедансу, виміряних в темноті та при освітленні білим світлом.

Результати. Встановлено, що срібло переважно осаджується у приповерхневому шарі поруватого кремнію, а його кількість залежить від тривалості електрохімічного впровадження. Легування поруватого шару сріблом зумовлює збільшення електропровідності та ємності структур на основі поруватого кремнію та відновленого оксиду графену. Опромінення поверхні експериментальних структур білим світлом спричиняє збільшення прямого і зворотного струму через структури, але майже не впливає на латеральні вольт-амперні характеристики через шунтування плівки відновленого оксиду графену наночастинами срібла. На основі аналізу спектрів імпедансу у частотному діапазоні 10^2 – 10^5 Гц досліджено вплив опромінення на резистивно-ємнісні властивості гібридних структур.

Висновки. Встановлено, що інкорпороване срібло утворює додаткові шляхи проходження струму через поруватий кремній і частково шунтує плівку відновленого оксиду графену. Як наслідок, отримані структури на основі поруватого кремнію та відновленого оксиду графену демонструють збільшення електропровідності при

перенесенні заряду як через структуру, так і вздовж поверхні. Крім того, наночастинки срібла сприяють накопиченню фотогенерованих носіїв заряду та їх виведенню з поруватого шару.

Ключові слова: Поруватий кремній, відновлений оксид графену, наночастинки срібла, фоточутливість, імпеданс.

УДК 538.93

ЗМІНИ ТЕРМОСТИМУЛЬОВАНОЇ ПРОВІДНОСТІ КРИСТАЛІВ CsPbBr_3 ТА CsPbCl_3 ПІДДАНИХ ДІЇ Х-ОПРОМІНЕННЯ

Богдан Павлик¹, Роман Лис¹, Роман Гамерник²,
Микола Гойдало¹

Львівський національний університет імені Івана Франка.

¹Кафедра сенсорної та напівпровідникової електроніки,
вул. Ген. Тарнавського, 107. Львів 79017, Україна.

²Кафедра експериментальної фізики,
вул. Кирила і Мефодія, 8. Львів 79005, Україна

Павлик Б.В. та ін. (2025). Зміни термостимульованої провідності кристалів CsPbBr_3 та CsPbCl_3 підданих дії Х-опромінення. *Електроніка та інформаційні технології*, 29, 145-154.
<https://doi.org/10.30970/eli.29.13>

АНОТАЦІЯ

Вступ. Кристали CsPbBr_3 та CsPbCl_3 залишаються одними із найбільш інтенсивно досліджуваних матеріалів. Це зумовлено їхнім високим коефіцієнтом корисної дії в сонячних батареях, яка зараз перевищує 22 %. Однак невирішеною проблемою є деградація даних перовскітів під впливом зовнішніх чинників.

Методика досліджень. В даній роботі досліджувалися монокристали перовскітів CsPbBr_3 та CsPbCl_3 , які були піддані впливу Х-опромінення. Експеримент спрямований на вивчення закономірностей дозних (Х-опромінення) змін електрофізичних характеристик кристалів. Таке дослідження є актуальним через високу чутливість матеріалів до УФ, Х-, та γ-випромінювання, що робить їх перспективними для виготовлення фотодетекторів в сфері виявлення іонізуючих випромінювань.

Результати досліджень. Виявлено, що зміни енергій активації електропровідності можуть бути зумовлені зміною механізму провідності внаслідок формування перовскітних фазових перетворень у кристалах CsPbBr_3 . Спостерігалось зменшення енергій активації провідності в області високих температур (160–180) °С, зумовлене дією Х-променів. Виявлена тенденція до зменшення енергій активації провідності при збільшенні величини дози опромінення. В області низьких температур (30–90) °С у кристалах CsPbBr_3 спостерігалися різкі зміни сили струму через зразки. Величина струму суттєво знижується в області (95–140) °С, і продовжується подальшим збільшенням в діапазоні вищих температур. Висловлено припущення, що такі результати можуть бути зумовлені генеруванням радіаційно-стимульованих дефектів вакансійного типу (V_{Br}). Утворення аналогічних фазових структур у кристалах CsPbCl_3 відбувається за значно нижчих температур і значно менших величин відповідних термострумів, тому для їхнього аналізу необхідні додаткові детальніші дослідження. Генерація радіаційних дефектів для кристалів CsPbCl_3 стає суттєвою за доз ≥ 130 Гр.

Висновки. У роботі встановлено, що дія Х-випромінювання і подальше проходження струму через кристал CsPbBr_3 супроводжується формуванням двох перовскітних фаз CsPb_2Br_5 та Cs_4PbBr_6 . Формування відповідних фазових структур у кристалах CsPbCl_3 відбувається за значно нижчих температур і значно менших величин відповідних термострумів, тому для їхнього аналізу необхідні додаткові детальніші дослідження.



© 2025 Bohdan Pavlyk et al. Published by the Ivan Franko National University of Lviv on behalf of Електроніка та інформаційні технології / Electronics and Information Technologies. This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Ключові слова: Х-опромінення, радіаційні дефекти, перовскіти, енергія активація провідності.

ВСТУП

Перовскітні матеріали стали найбільш перспективними та ефективними недорогими енергетичними матеріалами для різноманітних оптоелектронних і фотонних пристроїв. Унікальні фізичні властивості цих структур [1], такі як високий коефіцієнт поглинання, амбіполярний транспорт заряду на великі відстані, низька енергія зв'язку екситонів, висока діелектрична проникність, сегнетоелектричні властивості тощо, викликали величезний інтерес до цих матеріалів для оптоелектронних і фотоелектричних застосувань. Чудові сегнетоелектричні та надпровідні властивості перовскітів на основі оксидів були ретельно вивчені для різних застосувань [2]. Завдяки екситонним властивостям шаруватих перовскітних матеріалів (органіко-неорганічних галогенідів свинцю) ці матеріали були вивчені для застосування в тонкоплівкових світлодіодах (LED) і польових транзисторах (FET) [3]. Перовскітні матеріали також широко використовуються в світлодіодах, фотодетекторах, нанолазерах і хвилеводах [4].

Галогенідні напівпровідники свинцю з кристалічною структурою перовскіту та стехіометрією CsPbX_3 (де $X = \text{I}^-, \text{Br}^-, \text{Cl}^-$ або їх суміші), вирощені в розчині чи за допомогою метода Бріджмена [5], останнім часом стали, можливо, найбільш інтенсивно досліджуваним класом неорганічних оптоелектронних матеріалів, після демонстрації безпрецедентної продуктивності в якості поглиначів випромінювання із ефективністю перетворення електроенергії, яка зараз перевищує 22 %. Ці матеріали незабаром також можуть використовуватися у світловипромінюючих діодах, лазерах, ультрафіолетових та інфрачервоних фотодетекторах, а також в Х-детекторах та детекторах гамма-променів.

Таке різноманіття застосувань значною мірою пояснюється унікальними параметрами та характеристиками, високою дефектостійкістю цих напівпровідників: низька густина електронних пасток, незважаючи на велику концентрацію точкових дефектів. Кілька загальноприйнятих параметрів є прикладом оригінальної фотофізичної та електронної якості перовскітів: низька концентрація носіїв (10^9 – 10^{11} см^{-3}), низька густина пасток (10^9 – 10^{10} см^{-3}), які нижчі, ніж у монокристалічного Si, висока рухливість носіїв ($2,5$ – $1000 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$), тривалий час життя носіїв заряду ($0,08$ – 450 нс), великі електронно-діркові дифузійні довжини пробігу (2 – 175 мкм), малі ефективні маси носіїв заряду ($0,069$ – $0,25 m_0$), високі коефіцієнти оптичного поглинання на краю поглинання (1 – $4,5 \cdot 10^4 \text{ см}^{-1}$) і висока ефективність люмінесценції [5].

Електропровідність напівпровідникових сполук залежить від концентрації дефектів в кристалі і від температури. З підвищенням температури електропровідність напівпровідників збільшується, що зумовлено зростанням концентрації носіїв заряду. Низка досліджень [1–5], які розглядають високу іонну провідність в монокристалах цих матеріалів, пояснюють це міграцією аніонів. Інше дослідження [6] також підтверджує це, припускаючи, що ці матеріали є галогід-іонними провідниками та що їхня іонна провідність близька до інших добре відомих галогенід-іонних провідників PbBr_2 та PbCl_2 . Розрахована в цьому дослідженні енергія активації міграції становила $0,25 \text{ eV}$ для CsPbBr_3 . Зокрема, це дослідження пояснює, що провідність зумовлена міграціями аніонів [6].

Сполуки зі структурою перовскіту легко змінюють свою кристалічну структуру від ромбічної до кубічної через тетрагональну. Крім того, в певних діапазонах температур часто спостерігається співіснування кількох фаз. Фазове перетворення від ромбічної

до кубічної відбувається в діапазоні температур від кімнатної температури до 473 К. Це спостереження підтверджується дослідженнями [6] на цю тему, в яких також описані інші фазові перетворення, а саме тетрагональна фаза виділяється між ромбічною і кубічною.

Перовскітні плівки часто мають велику кількість внутрішніх дефектів, таких як антивузлові дефекти (тобто іони А на В-вузлах), міжвузлові дефекти та вакансії, а також домішки та висячі зв'язки на границях зерен і поверхнях, що може призвести до енергетичних станів, які значно сприяють безвипромінювальній рекомбінації фотоактивованих носіїв. На рівні пристрою така небажана рекомбінація неминуче збільшує дефіцит напруги холостого ходу, таким чином діючи як важлива перешкода на шляху до досягнення теоретичної межі ефективності [7].

Окрім термічного розкладання летких органічних катіонів, деградація під впливом вологи, кисню та світла також значно скорочує термін служби, і наявність дефектів, ймовірно є посередником цієї деградації [8-10].

Крім того, небажаний фазовий перехід FAPbI_3 і CsPbI_3 із перовскітної фази в неперовскітну фазу швидко перетворює верхній шар поглинання світла в оптичний мертвий шар протягом декількох днів [11-13]. Нарешті, погана відтворюваність і гістерезис залежності струм-напруга, які пов'язані з утворенням дефектів і міграцією, залишаються невирішеною проблемою [14, 15]. Підсумовуючи, можна стверджувати, що дефекти перовскіту відповідають за більшість проблемних питань, які терміново необхідно розв'язати для остаточної комерціалізації приладів на основі перовскітних кристалів.

МЕТОДИКА ДОСЛІДЖЕНЬ

У експерименті використовувались монокристали CsPbBr_3 та CsPbCl_3 вирощені методом Бріджмена, товщиною 2,3 мм із припаяними за допомогою срібної пасти контактами. Для визначення впливу іонізуючого випромінювання на електрофізичні властивості кристалів було досліджено температурні залежності зміни сили струму опромінених Х-променями (Х-променева установка УРС-1.0; мідний антикатод; струм пучка 10 мА; напруга 45 кВ) дозами 13, 43 та 130 Гр.

Температурні залежності струму через зразок досліджувалися у затемненому кристалі за постійної напруги 1 В, зразок при цьому не забарвлювався. Для вимірювання електропровідності кристалів контакти формувалися за допомогою срібної пасти. Товщина зразка вибиралася такою, щоб було повне поглинання Х-випромінювання, яка падало на нього.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

В неопроміненому кристалі (рис. 1, крива 1) можна відзначити слабку зміну значень струму, в діапазоні температур 20–90 °С, від фонового значення до $2 \cdot 10^{-8}$ А. Дана залежність зберігається при короткій (13 Гр) експозиції кристалу. При збільшенні дози опромінення (43–130 Гр) діапазон слабкої зміни розширюється до 115 °С (рис. 1, криві 3 та 4).

З врахуванням аналізу літературних даних [4] можна стверджувати, що вихідні кристали CsPbBr_3 володіють монофазністю в діапазоні температур від 20 до 90 °С і дозах опромінення до 13 Гр. Монофазність порушується, якщо кристал був підданий дії Х-випромінювання дозами більшими за 13 Гр (рис. 1, криві 3 та 4) і подальшого прикладання постійного струму через зразок.

Енергії активації термостимульованої електропровідності неопромінених і опромінених низькими дозами (до 13 Гр) кристалів CsPbBr_3 у низько- та високотемпературних перовскітних фазах практично співпадають між собою і

становлять 0,43 еВ. За дози опромінення $D > 43$ Гр енергія активації низькотемпературної області термостимульованої електропровідності зростає до величини 0,68 еВ, а в області високотемпературної фази вона зменшується до 0,36 еВ.

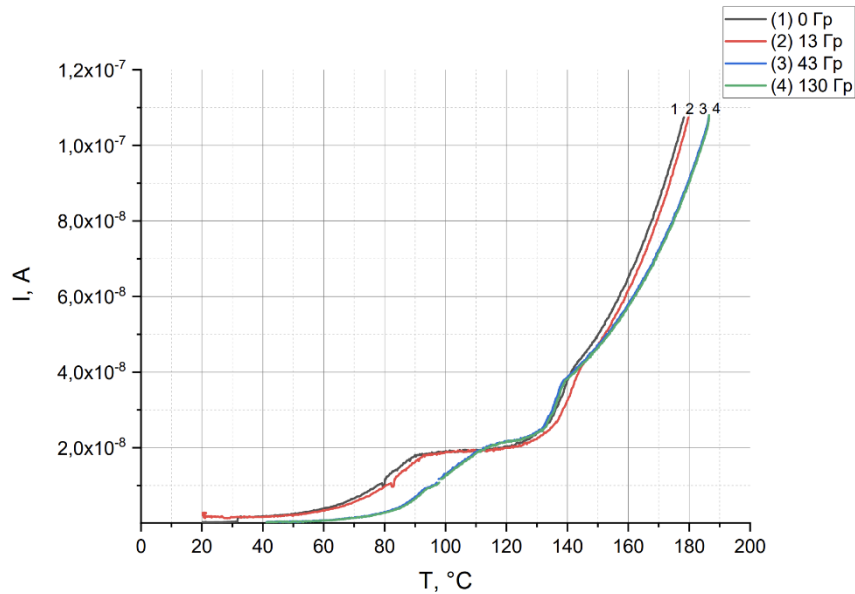
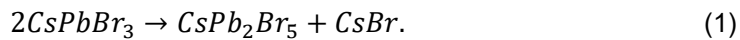
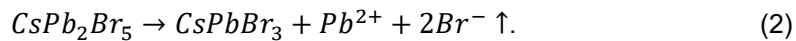


Рис. 1. X-променева зміна температурної залежності сили струму в кристалі CsPbBr_3 . $U = \text{const} = 1$ В.
Fig. 1. X-ray change of temperature dependence of current strength in a CsPbBr_3 crystal. $U = \text{const} = 1$ V.

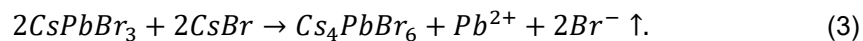
Можна припустити, що дія X-випромінювання, аналогічно як і дія атмосферної вологи [16], приводить до деградації монофазності кристала CsPbBr_3 з утворенням тетрагональних фаз CsPb_2Br_5 , які представляються реакцією:



Проходження струму I через зразок супроводжується наявністю іонної провідності завдяки іонам Br^- , а також рухом основних носіїв заряду – дірок. В кінцевому варіанті, коли іони бромю досягнуть анода, атоми бромю виділяються із зразка:



Надлишок CsBr у кристалі сприяє також утворенню фази Cs_4PbBr_6 [17] за схемою:



А це означає, що дія X-випромінювання і подальше проходження струму через кристал CsPbBr_3 за температури, меншої за 115 °C, супроводжується формуванням двох нових перовскітних фаз CsPb_2Br_5 та Cs_4PbBr_6 .

Внаслідок реакцій (2) та (3) можливе утворення атомів свинцю, які забарвлюють кристал CsPbBr_3 і зменшують його прозорість.

За подальшого підвищення температури, починаючи з 135°C , спостерігається різке збільшення величини струму через зразок (при $130^\circ\text{C} - 2 \cdot 10^{-8} \text{ A}$, а при $180^\circ\text{C} - 6 \cdot 10^{-8} \text{ A}$ у неопромінену та опромінену малою дозою ($D < 13 \text{ Gr}$) та $9 \cdot 10^{-8} \text{ A}$ в опромінених ($D > 43 \text{ Gr}$) кристалах). Як видно з графічної залежності (рис. 1) величини термостимульованих струмів через опромінений ($D > 43 \text{ Gr}$) зразок CsPbBr_3 зменшуються зі збільшенням температури. Стрімке збільшення величини термоструму за нагрівання зумовлене формуванням високотемпературної перовскітної фази в околі температур 130°C .

Зі збільшенням величини дози опромінення ($D > 13 \text{ Gr}$) в діапазоні температур $135\text{--}180^\circ\text{C}$ спостерігається незначне зменшення величини струму через зразок, що може бути зумовлено і генерацією радіаційних дефектів вакансійного типу, які є пастками для основних носіїв заряду.

На рис. 2 представлено залежності зміни характеру проходження термоструму через кристал CsPbCl_3 . Вимірювання електропровідності проводилося у затемненому кріостаті. Форми відповідних кривих (в залежності від величини дози опромінення) відрізняються від відповідних залежностей отриманих при дослідженні кристалів CsPbBr_3 , але добре корелюють із експериментальними результатами, представленими у роботі [18].

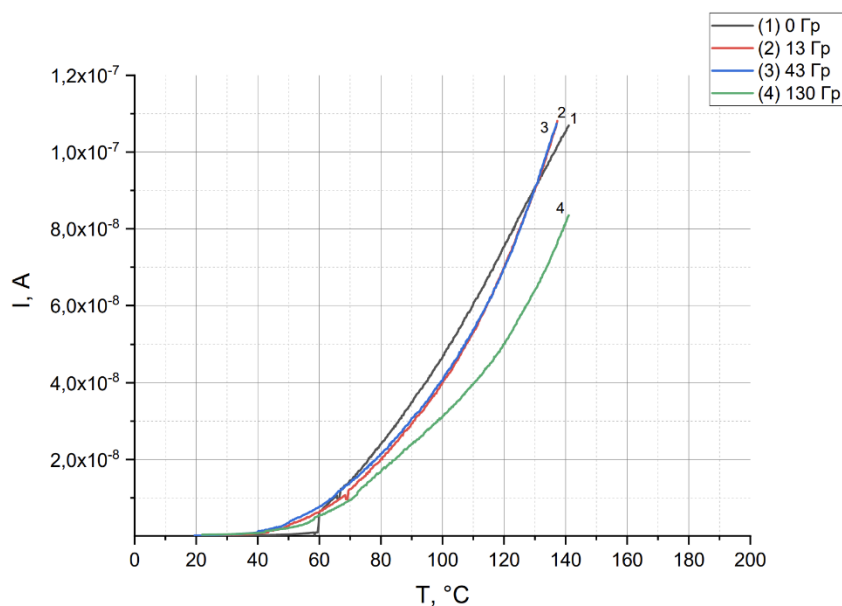


Рис. 2. Термостимульовані закономірності змін величини струму в опроміненому кристалі CsPbCl_3 . $U = \text{const} = 1 \text{ B}$.

Fig. 2. Thermally stimulated patterns of changes in the current magnitude in an irradiated CsPbCl_3 crystal. $U = \text{const} = 1 \text{ V}$.

Енергія активації термостимульованої електропровідності кристалів CsPbCl_3 як вихідних зразків, так і низькодозно ($D < 43 \text{ Gr}$) опромінених X-випромінюванням, є дещо більша, ніж у випадку кристалів CsPbBr_3 . Але зі збільшенням величини

поглинутої дози до 130 Гр, енергія активації, як і для кристалів CsPbBr_3 , зменшується до значення 0,49 еВ.

Оскільки формування відповідних фазових структур у кристалах CsPbCl_3 відбувається за значно нижчих температур і значно менших величин відповідних термострумів, тому для їх аналізу потрібні додаткові експериментальні дослідження. Звертає на себе той факт, що дія Х-опромінення ($D = 13$ Гр та $D = 43$ Гр) практично однаково проявляється і слабо відрізняються ці криві (рис. 2, криві 2 та 3) від кривої у випадку неопроміненого зразка (рис. 2, крива 1). Тільки за дози опромінення 130 Гр спостерігається більш стрімке зменшення величини термоструму (рис. 2, крива 4).

В області високотемпературної перовскітної фази CsPbCl_3 ($t = 140^\circ\text{C}$) величина струму зменшилась з $1,1 \cdot 10^{-7}$ А до величини $8,0 \cdot 10^{-8}$ А, що може бути зумовлено генерацією радіаційних дефектів, подібно як і у випадку з кристалами CsPbBr_3 .

Термостимульовані закономірності змін величини струму (рис. 3) через кристал CsPbBr_3 при охолодженні корелюють із відповідними даними, отриманими при їх нагріванні. Спостерігається тільки зменшення області температур існування перовскітних фазових перетворень: $90\text{--}135^\circ\text{C}$ при нагріванні та $120\text{--}130^\circ\text{C}$ при охолодженні.

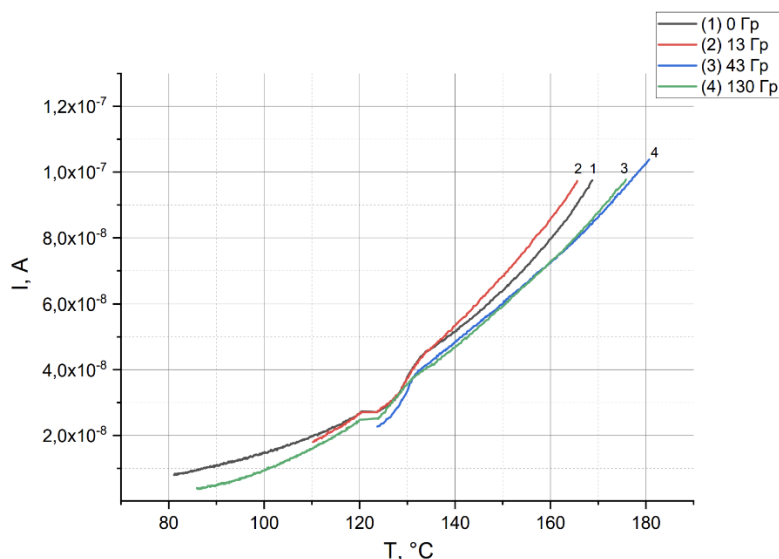


Рис. 3. Термостимульовані закономірності змін величини струму в опроміненому кристалі CsPbBr_3 при охолодженні. $U = \text{const} = 1$ В.

Fig. 3. Thermally stimulated regularities of changes in the current magnitude in an irradiated CsPbBr_3 crystal upon cooling. $U = \text{const} = 1$ V.

У випадку вимірювання закономірностей зміни величини струму через кристал CsPbCl_3 при його охолодженні (рис. 4) спостерігається кореляція величини струмів опромінених і неопромінених зразків тільки у високотемпературній ($t > 140^\circ\text{C}$) області. При охолодженні зразків опромінених дозами 130 Гр, в області температур менших 110°C , спостерігається більш швидке зменшення величини струму через зразок: при 90°C у неопромінених та опромінених ($D < 130$ Гр) – $I = (4,0\text{--}4,6) \cdot 10^{-8}$ А, а у випадку охолодження зразка опроміненого дозою 130 Гр – $I = 1,8 \cdot 10^{-8}$ А (рис. 4).

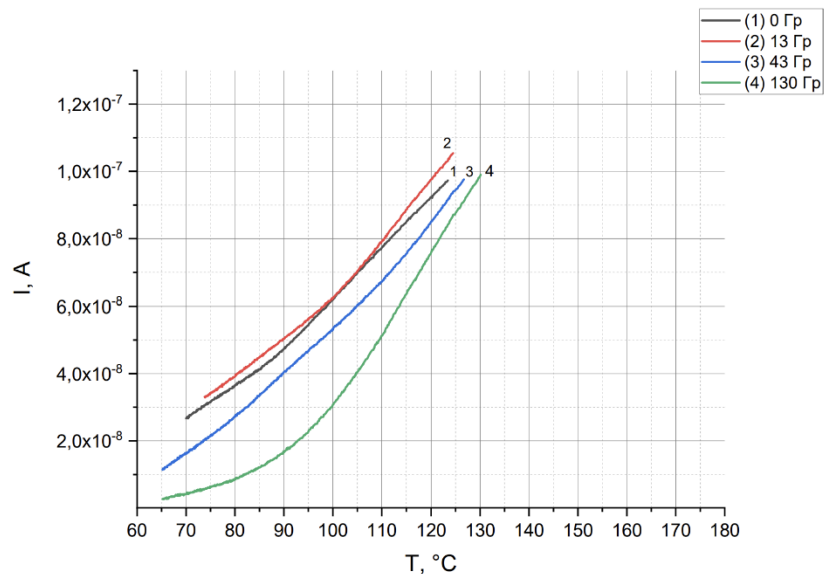


Рис. 4. Термостимульовані закономірності змін величини струму в опромінену кристалі CsPbCl_3 при охолодженні. $U = \text{const} = 1 \text{ В}$.

Fig. 4. Thermostimulated regularities of the current magnitude changes in the irradiated CsPbCl_3 crystal during cooling. $U = \text{const} = 1 \text{ V}$.

ВИСНОВКИ

Аналіз отриманих результатів дозволяє підтвердити зроблені припущення про співіснування в матриці кристалів CsPbBr_3 таких перовскітних фаз: CsPb_2Br_5 та Cs_4PbBr_6 . Ці фази можуть утворюватися за дії на кристал CsPbBr_3 іонізуючого Х-випромінювання за поглинутої дози $D \geq 43 \text{ Гр}$. Механізм перебудови перовскітних структур за дії на них радіації можна пояснити радіаційно-стимульованим вакансійним дефектоутворенням.

Формування відповідних фазових структур, аналогічних до структур для кристалів CsPbBr_3 , у кристалах CsPbCl_3 відбувається за значно нижчих температур і значно менших величин відповідних термострумів, тому для їхнього аналізу необхідні додаткові детальніші дослідження. Генерація радіаційних дефектів для кристалів CsPbCl_3 стає суттєвою за доз, не менших 130 Гр .

ДОТРИМАННЯ ЕТИЧНИХ СТАНДАРТІВ

Автори заявляють, що не мають конкуруючих інтересів.

АВТОРСЬКИЙ ВНЕСОК

Загальна концептуалізація [Б.П., Р.Л., Р.Г., М.Г.]; методика [Б.П., Р.Л., Р.Г.]; дослідження [Б.П., Р.Л., Р.Г., М.Г.]; написання початкової версії статті [Б.П., Р.Л.]; написання кінцевої версії статті та редагування [Р.Г., М.Г.]; підготовка графічного матеріалу [Р.Л., М.Г.].

Усі автори ознайомилися та погодилися з опублікованою версією статті.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Jinsong Huang, Yongbo Yuan, Yuchuan Shao, Yanfa Yan. (2017). Understanding the physical properties of hybrid perovskites for photovoltaic applications. *Nature Reviews Materials*, 2(7), 17042 (1–19). <https://doi.org/10.1038/natrevmats.2017.42>
- [2] Pankaj P. Khirade, Anil V. Raut. (2022) Perovskite Structured Materials: Synthesis, Structure, Physical Properties and Applications. *Recent Advances in Multifunctional Perovskite Materials*, (1–31). <https://doi.org/10.5772/intechopen.106252>
- [3] Gnanasampanthan Abiram, Murugathas Thanihaichelvan, Punniamoorthy Ravirajan, Dhayalan Velauthapillai. (2022). Review on Perovskite Semiconductor Field–Effect Transistors and Their Applications. *Nanomaterials*, 12(14), 2396 (1–33). <https://doi.org/10.3390/nano12142396>
- [4] Reshmi Varma, P.C. (2018). Low-Dimensional Perovskite. *Perovskite Photovoltaics. Chapter 7 – Low-Dimensional Perovskites*, 197–229. <https://doi.org/10.1016/B978-0-12-812915-9.00007-1>
- [5] Dirin, D.N., Cherniukh, I., Yakunin, S., Shynkarenko, Y., Kovalenko, M.V. (2016). Solution-Grown CsPbBr₃ Perovskite Single Crystals for Photon Detection. *Chemistry of Materials*, 28, 8470–8474. <https://doi.org/10.1021/acs.chemmater.6b04298>
- [6] Kayun, I., Lys, R., Tymkiv, Y. (2023). Modeling of ion transfer processes in CsPbBr₃ crystals. *Proceedings of the 2023 IEEE 13th International conference on electronics and information technologies (ELIT 2023)*, 340–343. <https://doi.org/10.1109/ELIT61488.2023.10311009>
- [7] Zhendong Guo, Man Yuan, Gaoyuan Chen, Feng Liu, Ruifeng Lu, Wan-Jian Yin. (2024). Understanding Defects in Perovskite Solar Cells through Computation: Current Knowledge and Future Challenge. *Advanced Science*, 11(20), 2305799(1–13). <https://doi.org/10.1002/advs.202305799>
- [8] Aurélien M.A. Leguy, Yinghong Hu, Mariano Campoy-Quiles, M. Isabel Alonso, Oliver J. Weber, Pooya Azarhoosh, Mark van Schilfgaarde, Mark T. Weller, Thomas Bein, Jenny Nelson, Pablo Docampo, Piers R. F. Barnes (2015). Reversible Hydration of CH₃NH₃PbI₃ in Films, Single Crystals, and Solar Cells. *Chemistry of Materials*, 27(9), 3397–3407. <https://doi.org/10.1021/acs.chemmater.5b00660>
- [9] Namyoung Ahn, Kwisung Kwak, Min Seok Jang, Heetae Yoon, Byung Yang Lee, Jong-Kwon Lee, Peter V. Pikhitsa, Junseop Byun, Mansoo Choi. (2016). Trapped charge-driven degradation of perovskite solar cells. *Nature Communications*, 7(13422), 1–9. <https://doi.org/10.1038/ncomms13422>
- [10] Nicholas Aristidou, Christopher Eames, Irene Sanchez-Molina, Xiangnan Bu, Jan Kosco, M. Saiful Islam, Saif A. Haque. (2017). Fast oxygen diffusion and iodide defects mediate oxygen-induced degradation of perovskite solar cells. *Nature Communications*, 8(15218), 1–10. <https://doi.org/10.1038/ncomms15218>
- [11] Wei Hui, Lingfeng Chao, Hui Lu, Fei Xia, Qi Wei, Zhenhuang Su, Tingting Niu, Lei Tao, Bin Du, Deli Li, Yue Wang, He Dong, Shouwei Zuo, Bixin Li, Wei Shi, Xueqin Ran, Ping Li, Hui Zhang, Zhongbin Wu, Chenxin Ran, Lin Song, Guichuan Xing, Xingyu Gao, Jing Zhang, Yingdong Xia, Yonghua Chen and Wei Huang. (2021). Stabilizing black-phase formamidinium perovskite formation at room temperature and high humidity. *Science*, 371(6536), 1359–1364. <https://doi.org/10.1126/science.abf7652>
- [12] Shaun Tan, Ilhan Yavuz, Marc H. Weber, Tianyi Huang, Chung-Hao Chen, Rui Wang, Hao-Cheng Wang, Jeong Hoon Ko, Selbi Nuryyeva, Jingjing Xue, Yepin Zhao, Kung-Hwa Wei, Jin-Wook Lee, Yang Yang. (2020). Shallow Iodine Defects Accelerate the Degradation of α -Phase Formamidinium Perovskite. *Joule*, 4(11), 2426–2442. <https://doi.org/10.1016/j.joule.2020.08.016>

- [13] Yong Wang, Yuetian Chen, Taiyang Zhang, Xingtao Wang, Yixin Zhao. (2020). Chemically Stable Black Phase CsPbI₃ Inorganic Perovskites for High-Efficiency Photovoltaics. *Advanced Materials*, 32(45), 2001025. <https://doi.org/10.1002/adma.202001025>
 - [14] Dane W. de Quilletes, Sarah M. Vorpahl, Samuel D. Stranks, Hirokazu Nagaoka, Giles E. Eperon, Mark E. Ziffer, Henry J. Snaith and David S. Ginger. (2015). Solar cells. Impact of microstructure on local carrier lifetime in perovskite solar cells. *Science*, 348(6235), 683-6. <https://doi.org/10.1126/science.aaa5333>
 - [15] Jie Xing, Qi Wang, Qingfeng Dong, Yongbo Yuan, Yanjun Fang and Jinsong Huang. (2016). Ultrafast ion migration in hybrid perovskite polycrystalline thin films under light and suppression in single crystals. *Physical Chemistry Chemical Physics*, 18, 30484–30490, <https://doi.org/10.1039/C6CP06496E>
 - [16] Turedi, B., Lee, K.J., Dursun, I., Alamer, B., Wu, Z., Alarousu, E., Mohammed, O.F., Cho, N., Bakr, O.M. (2018). Water-Induced Dimensionality Reduction in Metal-Halide Perovskites. *The Journal of Physical Chemistry C*, 122, 14128–14134. <https://doi.org/10.1021/acs.jpcc.8b01343>
 - [17] Zhang, Z., Zhu, Y., Wang, W., Zheng, W., Lin, R., Li, X., Zhang, H., Zhong, D., Huang, F. (2018). Aqueous Solution Growth of Millimeter-Sized Nongreen-Luminescent Wide Bandgap Cs₄PbBr₆ Bulk Crystal. *Crystal Growth & Defect*, 18, 6393–6398. <https://doi.org/10.1021/acs.cgd.8b00817>
 - [18] Bruzzi, M., Latino, M., Falsini, N., Calisi, N., Vinattieri, A. (2022). Electrical and Optical Characterization of CsPbCl₃ Films around the High-Temperature Phase Transitions. *Nanomaterials*, 12(3), 570(1–17). <https://doi.org/10.3390/nano12030570>
-

CHANGES IN THERMALLY STIMULATED CONDUCTIVITY OF CsPbBr₃ AND CsPbCl₃ CRYSTALS EXPOSED TO X-RAY IRRADIATION

¹Bohdan Pavlyk¹, Roman Lys¹, Roman Gamernyk², Mykola Goydalo¹

Ivan Franko National University of Lviv.

¹Department of Sensory and Semiconductor Electronics,
St. Tarnavskogo 107, Lviv 79017, Ukraine.

²Department of Experimental Physics,
St. Kyrylo and Mefodiy 8, Lviv 79005, Ukraine

ABSTRACT

Introduction. CsPbBr₃ and CsPbCl₃ crystals remain among the most intensively studied materials. This is due to their high efficiency in solar cells, which currently exceeds 22%. However, an unsolved problem is the degradation of these perovskites under the influence of external factors.

Materials and Methods. This study investigates monocrystals of perovskites CsPbBr₃ and CsPbCl₃ subjected to X-ray irradiation. The experiment aims to establish the dose-dependent patterns of changes in the electrical characteristics of the crystals. Research under these conditions is relevant due to the high photosensitivity of the materials in the UV, X-ray, and γ-ray spectra, making them promising for the fabrication of photodetectors in the detection of ionizing radiation.

Results and Discussion. It has been found that changes in the activation energies of electrical conductivity may result from alterations in the conduction mechanism due to the formation of perovskite phase transitions in CsPbBr_3 crystals. A decrease in activation energies of conductivity in the high-temperature range (160–180 °C) was observed, attributed to the effect of X-rays. A trend towards decreasing activation energies of conductivity with increasing dose of irradiation was identified. In the low-temperature range (30–90 °C), CsPbBr_3 crystals exhibited sharp changes in current through the samples, significantly decreasing in the range of (95–140) °C, followed by an increase in current at higher temperatures. It is suggested that these results may be attributed to radiation-stimulated generation of vacancy-type defects (V_{Br}) in CsPbBr_3 crystals. The formation of similar phase structures in CsPbCl_3 crystals occurs at significantly lower temperatures and much lower corresponding thermoelectric currents, necessitating further detailed investigations for their analysis. The generation of radiation defects in CsPbCl_3 crystals becomes significant at doses ≥ 130 Gy.

Conclusions. The study establishes that the influence of X-ray irradiation and the subsequent passage of current through CsPbBr_3 crystals is accompanied by the formation of two perovskite phases, CsPb_2Br_5 and Cs_4PbBr_6 . The formation of the corresponding phase structures in CsPbCl_3 crystals occurs at much lower temperatures and much smaller values of the corresponding thermal currents, therefore, additional, more detailed studies are required for their analysis.

Keywords: X-ray irradiation, radiation defects, perovskites, conduction activation energy.

Збірник наукових праць

Електроніка та інформаційні технології

Випуск 29

2025

Підп. до друку 28.03.2025. Формат 70х100,16. Папір друк.
Друк на різогр. Гарнітура Times New Roman. Умовн. друк. арк. .
Тираж 100 прим. Зам. № .

Львівський національний університет імені Івана Франка.
79000 Львів, вул. Університетська, 1.

Свідоцтво про внесення суб'єкта видавничої справи до Державного
реєстру видавців, виготівників і розповсюджувачів видавничої
продукції. Серія ДК № 3059 від 13.12.2007 р.