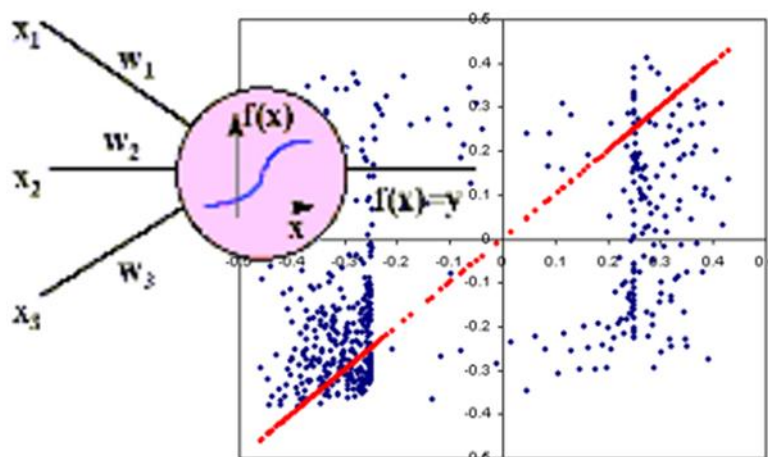


Ministry of Education and Science of Ukraine  
National Academy of Sciences of Ukraine  
National Academy of Sciences of Higher Education of Ukraine  
Institute of Artificial Intelligence Problems (Ukraine)  
Donbass State Engineering Academy (Ukraine)  
Vinnytsia National Technical University (Ukraine)  
Ternopil Ivan Puluj National Technical University (Ukraine)  
Engineering Academy of Serbia (IAS), Belgrade (Serbia)  
University of Szeged (Hungary)  
Apeiron University in Banja Luka (Bosnia and Herzegovina)  
DAAAM International Vienna  
Mechanical Engineering Faculty in Slavonski Brod, JJ Strossmayer University of Osijek  
(Croatia)  
University of Montenegro Faculty of Mechanical Engineering  
University of Zielona Góra (Poland)  
Vinnytsia National Agrarian University (Ukraine)  
Vinnytsia National Technical University (Ukraine)  
"American Journal Neural Network and Application" (USA)  
Problem Laboratory of Mobile Intelligent Technological Machines (Ukraine)



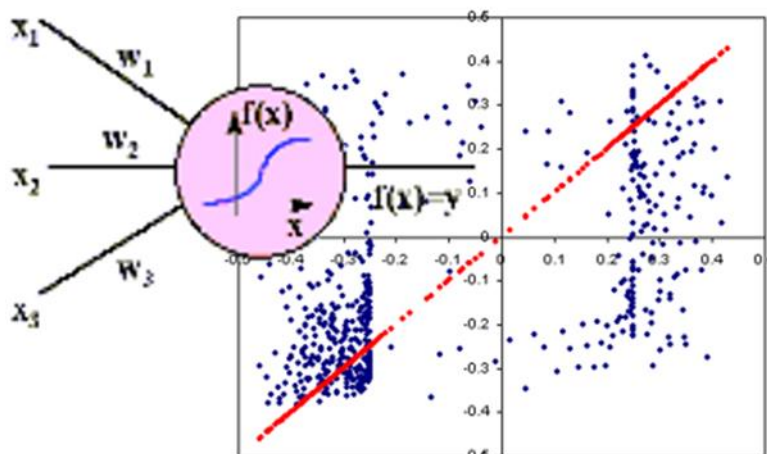
## COLLECTION OF SCIENTIFIC WORKS

XXIII International Scientific Conference

### «NEURAL NETWORK TECHNOLOGIES AND THEIR APPLICATIONS NNTA-2024»

Kramatorsk -Vinnytsia-Ternopil, Ukraine

Міністерство освіти і науки України  
Національна академія наук України  
Національна академія наук вищої освіти України  
Інститут проблем штучного інтелекту (Україна)  
Донбаська державна машинобудівна академія (Україна)  
Вінницький національний технічний університет (Україна)  
Тернопільський національний технічний університет імені Івана Пулюя (Україна)  
Engineering Academy of Serbia (IAS), Belgrade (Serbia)  
University of Szeged (Hungary)  
Apeiron University in Banja Luka (Bosnia and Herzegovina)  
DAAAM International Vienna  
Mechanical Engineering Faculty in Slavonski Brod, JJ Strossmayer University of Osijek  
(Croatia)  
University of Montenegro Faculty of Mechanical Engineering  
University of Zielona Góra (Poland)  
Вінницький національний аграрний університет (Україна)  
"American Journal Neural Network and Application" (США)  
Проблемна лабораторія мобільних інтелектуальних технологічних машин (Україна)



# **ЗБІРНИК НАУКОВИХ ПРАЦЬ**

## **XXIII Міжнародної наукової конференції**

### **«НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ НМТЗ-2024»**

Краматорськ-Вінниця-Тернопіль, Україна

**Рецензенти:**

Рамазанов С.К., докт.техн.наук, д.е.н.,професор, Київський національний університет імені Тараса Шевченка;

Суботін С.О., докт. техн. наук, професор, Запорізький національний технічний університет

Рекомендовано

вченою радою Донбаської державної машинобудівної академії  
(протокол № 4 від 28.11.2024)

**Програмний комітет конференції**

**Dašić Predrag** - Hon.D.Sc., Prof., Engineering Academy of Serbia (IAS), Belgrade,Serbia;

**Gyula Mester**, - Dr.Sc., Professor, Academic, University of Szeged, Hungary;

**Jenek Mariusz** - Dr. Inz, Universitet Zielonogorski, Poland;

**Marušić Vlatko** - Dr.Sc., Prof., J.J. Strossmayer University of Osijek, Mechanical Engineering Faculty in Slavonski Brod, Croatia;

**Guida Domenico** - Dr.Sc., Prof., University of Salerno, Department of Industrial Engineering (DIn), Fisciano, Italy;

**Karabegović Isak** - Dr.Sc., Prof., Academy of Sciences and Arts of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina;

**Mirjanić Dragoljub** - Dr.Sc., Prof., Academy of Sciences and Arts of the Republika Srpska (ANURS), Banja Luka, Republic of Srpska, Bosnia and Herzegovina;

**Nedeff Valentin** - Dr.Sc., Prof., University of Bacău, Faculty of Engineering, Bacău, Romania;

**Pele Alexandru-Viorel**, dean - Dr.Sc., Prof., University of Oradea, Faculty of Management and Technological Engineering, Oradea, Romania;

**Raul Turmanidze** - Georgian Technical University (GTU), Faculty of Transportation and Mechanical Engineering, Tbilisi, Georgia;

**Zhelezarov S. Piya**, rector - Dr.Sc., Prof., Technical University of Gabrovo, Gabrovo,Bulgaria;

**Гринь О.Г.** – к.т.н., доц., ДДМА, м. Краматорськ-Тернопіль, Україна;

**Дмитрієв Д.О.** – д.т.н., проф., ХНТУ, м. Херсон, Україна;

**Клименко С.А.** – д.т.н., проф. ІНМ АНУ, Київ, Україна;

**Ковалевська О.С.** - к.т.н. ,доц., ДДМА, м. Краматорськ, Україна;

**Ковалевський С.В.** - д.т.н., проф., ДДМА, м. Краматорськ, Україна;

**Ковальов В.Д.** - д.т.н., проф., ДДМА, м. Краматорськ, Україна;

**Козлов Л.Г.** - д.т.н., проф., ВНТУ, м. Вінниця, Україна;

**Кузнецов Ю.М.** - д.т.н., проф., НТУУ «КПІ» ім. І. Сікорського, м. Київ, Україна;

**Макаренко Н.О.** - д.т.н., проф., ДДМА, м. Краматорськ, Україна;

**Новіков Ф.В.** - д.т.н., проф., ХНЕУ, м. Харків, Україна;

**Петров О.В.**- к.т.н., доц., ВНТУ, м. Вінниця, Україна;

**Рамазанов С.К.** – д.т.н., д.е.н., проф., КНЕУ, м. Київ, Україна;

**Сапон С.П.** – к.т.н., доц., ЧНТУ, м. Чернігів Україна;

**Суботін С.О.** – д.т.н., проф., ЗНУ, м. Запоріжжя, Україна;

**Сухоруков С.І.** – к.т.н., доц. ВНТУ, м. Вінниця, Україна;

**Турчанін М.А.**- д.х.н., проф., ДДМА, м. Краматорськ, Україна.

Н46      Нейромережні технології та їх застосування НМТіЗ-2024: збірник наукових праць XXIII Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2024» / за заг. ред. д-ра техн. наук, проф. С.В.Ковалевського. – Краматорськ-Тернопіль: ДДМА, 2024. – 237 с.

ISBN 978-617-7889-91-4

У збірнику праць представлені перспективні теоретичні та практичні розробки в області нейромережних технологій, виконані в 2024 р. науковими школами України і світу. Розглядається можливість застосування нейронних мереж для управління об'єктами в режимі реального часу і особливості нейронного керування динамічними об'єктами. Наводиться ряд розробок по застосуванню нейронних мереж в різних областях практичної і науково-дослідної діяльності та створенню інтелектуальної системи для підвищення швидкості та зниження трудомісткості технологічної підготовки виготовлення нових виробів.

Для здобувачів освіти, наукових працівників широкого профілю та фахівців.

ISBN 978-617-7889-91-4

УДК 004.8:62-5:007

©ДДМА,2024

## ЗМІСТ

	Стор.
1. <b>Ковалевський С.В.</b> ( <i>Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль, Україна</i> ) <b>XXIII Міжнародна наукова конференція «НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ - НМТІЗ-2024».</b>	8
2. <b>Andrij Medvid, Vitaliy Yakovyna</b> ( <i>Lviv Polytechnic National University, Lviv, Ukraine</i> ) <b>ROBOT SELF COLLISION PREDICTION USING KOLMOGOROV-ARNOLD NETWORKS.</b>	11
3. <b>Hesheva Hanna</b> ( <i>Dmytro Motornyi Tavria State Agrotechnological University, Zaporizhzhya, Ukraine</i> ) <b>ANALYSIS OF THE EFFICIENCY OF NEURAL NETWORKS IN AUTONOMOUS CONTROL SYSTEMS COMPARED TO TRADITIONAL APPROACHES.</b>	17
4. <b>Kunichik O.V.</b> ( <i>Taras Shevchenko National University of Kyiv, Kyiv, Ukraine</i> ) <b>CLOUD-BASED LANDMINE DETECTION SERVICE WITH MESSENGER BOT INTEGRATION.</b>	21
5. <b>Olena Kovalevska</b> ( <i>Donbas State Engineering Academy, Kramatorsk - Ternopil, Ukraine</i> ) <b>ARTIFICIAL INTELLIGENCE AS A DRIVER OF TERRITORIAL COMMUNITY DEVELOPMENT.</b>	26
6. <b>Predrag Pravdić</b> ( <i>Academy of Professional Studies, Department in Kruševac, Serbia</i> ), <b>Violeta Đorđević, Jelena Erić-Obućina, Snežana Gavrilović</b> ( <i>Academy of Professional Studies, Department in Trstenik, Serbia</i> ) <b>INNOVATIVE APPROACHES TO LEARNING AND ADAPTIVE LEARNING.</b>	29
7. <b>Predrag Pravdić</b> ( <i>Academy of Professional Studies, Department in Kruševac, Serbia</i> ), <b>Violeta Đorđević, Jelena Erić-Obućina, Snežana Gavrilović</b> ( <i>Academy of Professional Studies, Department in Trstenik, Serbia</i> ) <b>PERSONALIZED EDUCATIONAL PROGRAMS WITH AI.</b>	38
8. <b>Sergiy Kovalevskyy</b> ( <i>Donbas State Engineering Academy, Kramatorsk - Ternopil, Ukraine</i> ), <b>Predrag Dasic</b> ( <i>Engineering Academy of Serbia (IAS), Belgrade, Serbia</i> ) <b>NEURAL NETWORK BASIS OF INNOVATION.</b>	46
9. <b>Vitaliy Yakovyna, Danylo Melnyk</b> ( <i>Lviv Polytechnic National University, Lviv Ukraine</i> ) <b>AN INTERPRETABLE SYSTEM FOR BREAST CANCER DIAGNOSIS USING XAI METHODS.</b>	51
10. <b>Афонін Д.І.</b> ( <i>Київський національний університет імені Тараса Шевченка, Київ, Україна</i> ) <b>МОДЕЛЮВАННЯ ПОШИРЕННЯ ІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ З АГЕНТАМИ.</b>	59
11. <b>Віт Р.В., Мазурець О.В.</b> ( <i>Хмельницький національний університет, м. Хмельницький, Україна</i> ) <b>ТЕМАТИЧНА КЛАСИФІКАЦІЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ.</b>	63
12. <b>Ємець К. В.</b> ( <i>Національний університет «Львівська політехніка», Львів, Україна</i> ) <b>ПІДХОДИ ДО ПОПЕРЕДНЬОГО ОПРАЦЮВАННЯ ЧАСОВИХ РЯДІВ ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ ПРОГНОЗУВАННЯ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ.</b>	67

13. **Жмай О. В.** ( *ПП «ОФСЕТІК», м. Одеса, Україна*) **ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ДІТЕЙ ТА ПІДЛІТКІВ: НАВЧАННЯ, РОЗВИТОК, БЛАГОПОЛУЧЧЯ.** 69
14. **Ковалевський С.В.** (*Донбаська державна машинобудівна академія, м. Краматорськ - Тернопіль, Україна*), **Козлов С.Л.** (*Вінницький національний технічний університет, м. Вінниця, Україна*) **ЗАСТОСУВАННЯ ДИФУЗНИХ МОДЕЛЕЙ ДО ВИРІШЕННЯ ЗАДАЧ SUPER-RESOLUTION.** 81
15. **Ланде Д. В., Гуменюк О. О.** (*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м.Київ, Україна*) **МЕТОДИКА ВИЗНАЧЕННЯ РІВНЯ ДОСТОВІРНОСТІ БІОГРАФІЙ З ВІКІПЕДІЇ НА БАЗІ СЕМАНТИЧНОГО НЕТВОРКІНГУ.** 84
16. **Мазурець О.В., Овчарук О.М.,** (*Хмельницький національний університет, м. Хмельницький, Україна*) **ДІАГНОСТУВАННЯ ПРОЯВІВ ПОСТТРАВМАТИЧНОГО СТРЕСОВОГО РОЗЛАДУ ЗА НЕЙРОМЕРЕЖЕВИМ АНАЛІЗОМ ТЕКСТОВОГО КОНТЕНТУ.** 93
17. **Мартинов А.Ю., Радюк П.М.** (*Хмельницький національний університет, м. Хмельницький, Україна*) **ШТУЧНИЙ ІНТЕЛЕКТ ЯК ОСНОВА ДЛЯ РОЗПІЗНАВАННЯ БПЛА ПО РАДІОСИГНАТУРІ.** 98
18. **Мельников О.Ю., Денисенко В. О.** (*Донбаська державна машинобудівна академія, м. Краматорськ - Тернопіль, Україна*) **ПРОГНОЗУВАННЯ ЗМІНИ РІВНЯ ЛІСОВОГО ВКРИТТЯ ОКРЕМОГО ЛІСНИЦТВА ЗА ДОПОМОГОЮ СУПУТНИКОВОГО С.ЕРВІСУ LANDSAT ТА ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ.** 104
19. **Молчанова М.О.** (*Хмельницький національний університет, м. Хмельницький, Україна*) **ВИЯВЛЕННЯ ОБ'ЄКТІВ ПРОПАГАНДИ У ТЕКСТОВИХ ПОВІДОМЛЕННЯХ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ ІЗ ВІЗУАЛЬНОЮ ІНТЕРПРЕТАЦІЄЮ РЕЗУЛЬТАТІВ.** 109
20. **Павлюк О.М., Міщук М.В., Заболотна А.О., Літовська О.В., Костінський Д.О.** (*Національний університет «Львівська політехніка», м. Львів, Україна*) **НЕЙРОМЕРЕЖНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ КОМПЛЕКСНОЇ ДІЯЛЬНОСТІ ПРОМИСЛОВОГО ПЕРСОНАЛУ ЗА ДАНИМИ СМАРТ-ГОДИННИКІВ.** 115
21. **Петров Д. Д.** (*Національний університет «Львівська політехніка», м. Львів, Україна*) **ПОКРАЩЕННЯ ЯКОСТІ ЗВ'ЯЗКУ GSM МЕРЕЖІ RL АГЕНТОМ.** 121
22. **Рамазанов С. К.,** (*Східноукраїнський національний університет імені Володимира Даля, м. Київ, Україна*) **КОГНІТИВНІ ТЕХНОЛОГІЇ ПРИЙНЯТТЯ ІНТЕЛЕКТУАЛЬНИХ РІШЕНЬ У НЕЙРОЕКОМОМІКІ У СКЛАДНОМУ ІНФОРМАЦІЙНОМУ СЕРЕДОВИЩІ.** 127

23. **Слюсар В.І.** (*Центральний науково-дослідний інститут озброєння та військової техніки ЗС України*) **НЕЙРОМЕРЕЖНИЙ АНАЛІЗ СТРУКТУРИ ХМАР НА БОРТУ АВІАЦІЙНИХ ПЛАТФОРМ ДЛЯ ПРОГНОЗУВАННЯ МЕТЕОДАНИХ У ПРИЗЕМНОМУ СЕРЕДОВИЩІ.** 137
24. **Собко О.В.,** (*Хмельницький національний університет, м. Хмельницький, Україна*) **ВІЗУАЛЬНА ІНТЕРПРЕТАЦІЯ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ КІБЕРБУЛІНГУ У ЦИФРОВИХ ТЕКСТАХ.** 146
25. **Чабан О.Р., Манзюк Е.А.** (*Хмельницький національний університет, м. Хмельницький, Україна*) **ПІДХІД ДО ІНТЕГРУВАННЯ ЕКСПЕРТНИХ ЗНАТЬ В МОДЕЛЬ U-NET ДЛЯ СЕГМЕНТУВАННЯ ЗОБРАЖЕНЬ МРТ СЕРЦЯ.** 153
26. **Anders T. Sandnes, Bjarne Grimstad, Odd Kolbjørnsen** (*Norwegian University of Science and Technology, Trondheim, Norway*) **MULTI-TASK LEARNING WITH LEARNED CONTEXTUAL INPUTS.** 158
27. **Andreas Rauscher, Johannes Kaiser, Manoj Devaraju, Christian Endisch** (*Technische Hochschule Ingolstadt, Institute of Innovative Mobility (IIMo), Research Team Electromobility and Learning Systems (ELS), Ingolstadt, Germany*) **DEEP LEARNING FOR PARTIAL DISCHARGE DETECTION IN ELECTRICAL MACHINES.** 161
28. **Antonello Cammarano, Vincenzo Varriale, Francesca Michelino, Mauro Caputo** (*University of Salerno, Fisciano, (SA), Italy*) **DISCOVERING TECHNOLOGICAL OPPORTUNITIES WITH NEURAL NETWORKS AND LITERATURE ANALYSIS.** 166
29. **Carmen Bisogni, Lucia Cimmino, Michele Nappi, Toni Pannese, Chiara Pero** (*University of Salerno, Department of Computer Science & Department of Management & Innovation Systems, Fisciano, Salerno, Italy*) **GAIT-BASED EMOTION RECOGNITION WITH PRIVACY PRESERVATION.** 170
30. **Daniel H. Stolfi, Grégoire Danoy** (*Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg*) **EVOLUTIONARY SWARM FORMATION FOR AUTONOMOUS ROBOTS.** 175
31. **Daniel Marek, Jakub Nalepa** (*Department of Algorithmics and Software, Silesian University of Technology, Gliwice, Poland; KP Labs, Gliwice, Poland*) **DEEP LEARNING FOR EXTRATERRESTRIAL ROCK SEGMENTATION.** 178
32. **Diego Echeverria-Rios, Peter L. Green** (*School of Engineering, The University of Liverpool, UK*) **GAUSSIAN PROCESSES FOR PREDICTING PRODUCT QUALITY IN MANUFACTURING.** 181
33. **Dominik Pieczyński, Bartosz Ptak, Marek Kraft, Mateusz Piechocki, Przemysław Aszkowski** (*Poznań University of Technology, Institute of Robotics and Machine Intelligence, Piotrowo 3A, Poznań, Poland*) **FAST, LIGHTWEIGHT DEEP LEARNING PIPELINE FOR UAV LANDING SUPPORT.** 184

34. **Federico Lombardo, Federico Pittino, Daniele Goldoni, Luca Selmi** 188  
(*University of Modena and Reggio Emilia, Modena, Italy*). **MACHINE LEARNING FOR NANOPARTICLE SIZING WITH BIOSENSOR ARRAYS.**
35. **Federico Rossi, Cinzia Bernardeschi, Marco Cococcioni** 192  
(*University of Pisa, Pisa, Italy*) **NEURAL NETWORK VERIFICATION IN CLOSED-LOOP SYSTEMS USING INTERVAL ARITHMETIC).**
36. **Gergő Békési, Lilla Barancsuk, Bálint Hartmann** 196  
(*Budapest University of Technology and Economics, Budapest, Hungary*) **DEEP NEURAL NETWORK STATE ESTIMATION WITH HYPERPARAMETER OPTIMIZATION.**
37. **Giampiero Bardella, Simone Franchini, Pierpaolo Pani, Stefano Ferraina** 199  
(*Sapienza University of Rome, Rome, Italy*) **LATTICE PHYSICS APPROACHES FOR NEURAL NETWORKS.**
38. **Lilli Frison, Simon Gölzhäuser, Moritz Bitterling, Wolfgang Kramer** 206  
(*Fraunhofer Institute for Solar Energy Systems ISE, Freiburg, Germany*) **ARTIFICIAL NEURAL NETWORKS FOR OPTIMIZING DISTRICT HEATING NETWORKS.**
39. **Łukasz Sobolewski** 211  
(*University of Zielona Góra, Zielona Góra, Poland*) **QUALITY EVALUATION OF POLISH TIMESCALE FORECASTING USING GMDH NEURAL NETWORK.**
40. **Mateusz Tejer, Rafal Szczepanski, Tomasz Tarczewski** 215  
(*Institute of Engineering and Technology, Nicolaus Copernicus University, Torun, Poland*) **ROBUST TASK SCHEDULING IN ROBOTICS USING REINFORCEMENT LEARNING.**
41. **Mathias Held, Jannis Bulling, Yevgeniya Lugovtsova, Jens Prager** 219  
(*Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany*) **ELASTIC CONSTANTS FROM ULTRASONIC DISPERSION IMAGES VIA NEURAL NETWORKS.**
42. **Paweł Majewski, Mariusz Mrzyglód, Piotr Lampa, Robert Burduk, Jacek Reiner** 223  
(*Wrocław University of Science and Technology, Wrocław, Poland*) **MONITORING INSECT LARVAE GROWTH WITH REGRESSION CNN AND KNOWLEDGE TRANSFER.**
43. **Tamara Sobot, Vladimir Stankovic, Lina Stankovic** 228  
(*University of Strathclyde, United Kingdom*) **HUMAN-IN-THE-LOOP ACTIVE LEARNING FOR TIME-SERIES ELECTRICAL DATA.**
44. **Tomaz Brzin, Miha Brojan** 234  
(*University of Ljubljana, Ljubljana, Slovenia*) **GAN-BASED INVERSE DESIGN OF SOFT MORPHING COMPOSITE BEAMS.**

**Ковалевський С.В.** (Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль, Україна)

## **XXIII Міжнародна наукова конференція «НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ - НМТІЗ-2024».**

### **Шановні колеги, учасники та гості конференції!**

Від імені організаційного комітету щиро вітаю вас на XXIII Міжнародній науковій конференції "Нейромережні технології та їх застосування". Це визначна подія у науковому та технічному світі, яка об'єднує провідних учених, дослідників, інженерів і практиків із різних куточків світу. Ваша присутність сьогодні є свідченням спільної мети — сприяти розвитку технологій майбутнього, які змінюють не лише науку та промисловість, але й наше повсякденне життя.

Нейромережні технології є одним із найперспективніших інструментів сучасності для реалізації інновацій у різних галузях економіки. Їх потенціал полягає у здатності кардинально змінювати підходи до вирішення стратегічних задач, забезпечуючи ефективність, точність і адаптивність. Сьогодні вони стають основою для автоматизації процесів, аналізу даних і прогнозування трендів, що є критично важливим у динамічному світі.

Автоматизація виробничих ліній також є важливим напрямом. Нейромережі оптимізують параметри процесів у реальному часі, забезпечуючи високу якість продукції та адаптуючись до змінних умов. Наприклад, методи навчання з підкріпленням дозволяють системам самостійно реагувати на зношеність обладнання чи зміну характеристик сировини. Використання згорткових нейромереж для аналізу зображень автоматизує контроль якості, виявляючи навіть найменші дефекти.

У маркетингу та управлінні ланцюгами постачання нейромережі забезпечують аналіз великих обсягів даних, що дозволяє прогнозувати поведінку споживачів, оптимізувати запаси й створювати персоналізовані пропозиції. Наприклад, моделі GPT і BERT створюють інтерактивний клієнтський досвід, підвищуючи лояльність і продажі. Для промислових підприємств це означає можливість швидкої адаптації до нових ринкових вимог.

Для України нейромережні технології є унікальною можливістю інтегрувати Індустрію 4.0, що базується на автоматизованих рішеннях і IoT, та підготуватися до Індустрії 5.0, яка акцентує увагу на гармонії людини й машин. Вони дозволяють створювати людиноцентричні рішення, як-от розумні міста чи автоматизовані системи охорони здоров'я, забезпечуючи екологічну сталість.

Відновлення промислового, освітнього та наукового потенціалу України також можливе завдяки нейромережам. У промисловості це сприяє модернізації обладнання й підвищенню ефективності. В освіті адаптивні платформи забезпечують підготовку фахівців для високотехнологічних галузей, а в науці великі дані прискорюють фундаментальні дослідження.

Таким чином, нейромережні технології стають ключовим фактором економічного зростання, трансформації промисловості та підвищення якості життя. Їх впровадження забезпечує фундамент для інновацій і відповідає викликам сучасності.

Конференція, яку ми сьогодні відкриваємо, присвячена обговоренню передових досліджень, інновацій та практичних рішень у сфері нейромережних технологій. Це унікальна



платформа, де поєднуються теоретичні основи, інженерні розробки та міждисциплінарний підхід.

Цьогорічний збірник наукових праць конференції вражає своєю різноманітністю. У ньому представлено більш 45 публікацій, які демонструють глибину і широту досліджень у сфері нейромереж. Ми також з гордістю можемо сказати, що конференція стала майданчиком для міжнародної співпраці. Учасники з більш ніж десяти країн діляться своїми досягненнями, досвідом і баченням майбутнього. Це не лише можливість презентувати свої роботи, але й знайти партнерів для спільних досліджень, започаткувати нові проєкти та сприяти формуванню глобальної мережі знань.

Сьогодні перед нами стоїть завдання не лише обговорити сучасний стан науки, але й окреслити шляхи її розвитку. Нехай ця конференція стане джерелом натхнення, нових ідей та рішень, які допоможуть створити технологічно досконале майбутнє.

Щиро бажаю всім учасникам плідної роботи, змістовних дискусій та нових наукових відкриттів. Разом ми формуємо наукову спільноту, здатну впливати на розвиток сучасного світу.

### **Dear Colleagues, Participants, and Guests of the Conference!**

On behalf of the Organizing Committee, I am honored to welcome you to the XXIII International Scientific Conference "Neural Network Technologies and Their Applications." This event represents a significant milestone in the scientific and technical community, bringing together leading scientists, researchers, engineers, and practitioners from around the globe. Your presence here today reflects our shared mission — to advance future technologies that transform not only science and industry but also our everyday lives.

Neural network technologies are among the most promising tools of our time for implementing innovations across various economic sectors. Their potential lies in their ability to fundamentally change approaches to solving strategic problems, ensuring efficiency, precision, and adaptability. Today, they form the foundation for process automation, data analysis, and trend forecasting, which are critically important in our rapidly evolving world.

Another important application area is the automation of production lines. Neural networks optimize process parameters in real-time, maintaining high product quality and adapting to changing conditions. For example, reinforcement learning methods allow systems to independently respond to equipment wear or changes in raw material properties. The use of convolutional neural networks for image analysis also automates quality control, detecting even the smallest defects.

In marketing and supply chain management, neural networks enable the analysis of large datasets from sources such as online sales or social media. This helps predict consumer behavior, optimize inventory, and create personalized offers. Models like GPT and BERT enhance the customer experience, boosting loyalty and sales. For industrial enterprises, this means the ability to quickly adapt production lines to new market demands.

For Ukraine, neural network technologies present a unique opportunity to integrate Industry 4.0, based on automated solutions and IoT, and prepare for Industry 5.0, which emphasizes harmony between humans and machines. They enable the creation of human-centered solutions, such as smart cities and automated healthcare systems, while ensuring ecological sustainability.

The restoration of Ukraine's industrial, educational, and scientific potential is also achievable through neural networks. In industry, they drive equipment modernization and productivity improvements. In education, adaptive platforms prepare specialists for high-tech fields, while in science, big data accelerates fundamental research.

Thus, neural network technologies are becoming a key factor in economic growth, industrial transformation, and improving quality of life. Their implementation lays the foundation for innovation and addresses contemporary challenges.

The conference we are inaugurating today is dedicated to discussing cutting-edge research, innovations, and practical solutions in the field of neural network technologies. It is a unique platform where theoretical foundations, engineering developments, and interdisciplinary approaches converge.

This year's conference proceedings are impressive in their diversity, featuring over 45 publications that demonstrate the depth and breadth of research in the field. We are also proud to note that this conference serves as a platform for international collaboration. Participants from more than ten countries are sharing their achievements, experiences, and visions for the future. This is not only an opportunity to present one's work but also to find partners for joint research, initiate new projects, and contribute to the formation of a global knowledge network.

Today, we are tasked not only with discussing the current state of science but also with outlining its future directions. May this conference be a source of inspiration, new ideas, and solutions that help create a technologically advanced future.

I sincerely wish all participants productive work, meaningful discussions, and groundbreaking scientific discoveries. Together, we are building a scientific community capable of influencing the development of the modern world.

**With respect,**  
**Sergiy Kovalevskyy**  
Chair of the Organizing Committee,  
Doctor of Technical Sciences, Professor  
Donbas State Engineering Academy

**Andriy Medvid, Vitaliy Yakovyna** (*Lviv Polytechnic National University, Lviv, Ukraine*)

## **ROBOT SELF COLLISION PREDICTION USING KOLMOGOROV-ARNOLD NETWORKS**

**Abstract:** Collision detection is essential for ensuring the safety of robotic manipulators, particularly when robotic arms operate within complex environments. Self-collision detection, involving contacts between various parts of a robotic arm or between the arm and its base, often represents a computational bottleneck for path planning. Conventional physics-based and geometric collision-checking approaches, such as those using PyBullet, offer high accuracy but can be computationally expensive, limiting their utility for real-time applications. To address this challenge, we present a novel approach employing Kolmogorov-Arnold Networks (KAN) to predict self-collisions for a robotic manipulator mounted on a mobile base. Using a dataset of 10 million randomly generated states, labeled as colliding or non-colliding via the PyBullet library, we trained KAN implementation with modified parameters, including a spline order of 2 and grid size of 4. Our model achieved over 98% prediction accuracy on a held-out test set while providing significant speed improvements when using batching. Specifically, the KAN model processed approximately 93,728 test states per second, which is more than 20 times faster than the PyBullet library. Key contributions of this work include the application of KAN for fast self-collision prediction, a comparison of its performance with fully-connected networks and conventional simulation-based methods. This study underscores the effectiveness of neural network-based approaches for balancing speed and accuracy in collision detection, offering a scalable solution for complex robotic applications like trajectory planning.

**Keywords:** robotic manipulators, Kolmogorov-Arnold Networks (KAN), collision detection, machine learning.

### **1. Introduction**

Collision detection is critical to ensure the safe operation of systems with robotic manipulators, especially when robotic arms interact with their environment. Self-collision can be defined as a situation where different parts of a robotic arm come into contact with each other or with the robot's base. Collision detection is the most time consuming part of the most robotic arm path planning methods. Conventional collision-checking algorithms often rely on physics simulations or geometric models, such as those implemented using libraries like PyBullet [2]. While these approaches provide high accuracy, they can be computationally expensive, especially when applied to large numbers of states. Real-time applications, such as robotic path planning, require faster and more scalable methods to estimate potential collisions while maintaining reasonable accuracy.

In recent years, machine learning techniques have been explored as an alternative to traditional physics-based simulation methods for predicting collisions. In this work Kolmogorov-Arnold Network (KAN) [1] was trained to predict collisions between arm, arm tool and robot base. Our choice of network architecture reasoned by the ability of KAN to approximate complex, nonlinear functions with high precision. In this study, we utilize an efficient KAN implementation [3] to predict self-collisions for a robotic manipulator mounted on a moving base.

To train the KAN model, we generated a dataset comprising 10 million random arm states sampled uniformly across joint limits. Each state was labeled as either colliding or non-colliding based on comprehensive checks using PyBullet simulations. The model's performance was evaluated on a held-out test set, achieving over 98% accuracy in predicting self-collisions. Furthermore, our approach demonstrated significant speed improvements when evaluating large batches of states compared to traditional methods.

The primary contributions of this paper are as follows:

- We introduce the application of KAN for efficient self-collision prediction of the robot.
- We compare the predictive performance and computational speed of our approach with physics-based simulation methods.

- We demonstrate the practical utility of using KAN for preliminary collision estimation, making it suitable for applications such as trajectory planning where rapid assessments of multiple states are necessary.

The remainder of this paper is structured as follows: Section 2 reviews related work in collision detection for robotic systems. Section 3 outlines our methodology, including data generation and the KAN architecture used. Section 4 presents the results of our experiments, while Section 5 discusses the implications and limitations of our approach. Finally, Section 6 concludes the paper with a summary of findings and potential future work.

## 2. Related Work

In recent years, neural networks have been increasingly applied to collision detection in robotic systems, offering promising alternatives to traditional model-based methods. This section reviews notable studies that have utilized neural network architectures for collision detection, summarizing their methodologies and key findings.

Sharkawy and Aspragathos (2018) developed a neural network-based approach for human-robot collision detection [4]. They designed a multi-output neural network trained using data from the coupled dynamics of a manipulator, both with and without external contacts. The network utilized intrinsic joint position and torque sensors to detect unwanted collisions and identify the specific collided link. This method demonstrated effective collision detection capabilities, enhancing safety in human-robot interactions.

Czubenko and Kowalczyk (2021) proposed a simple neural network architecture for collision detection in collaborative robots [5]. They implemented a virtual force and torque sensor using a neural network, which was tested on the CURA6 robot prototype. The study compared four different approaches: auto-regressive, recurrent neural network, convolutional long short-term memory, and mixed convolutional LSTM network (MC-LSTM). The MC-LSTM architecture achieved the highest effectiveness, with a mean absolute prediction error of approximately 22 Nm and a collision detection F1 score of 0.85.

Xu et al. (2020) presented a deep learning approach for collision detection in industrial collaborative robots [6]. They designed a deep neural network model to learn robot collision signals and recognize collision occurrences. This data-driven method unified feature extraction from high-dimensional signals and decision processes, demonstrating improved collision detection performance in industrial settings.

In the article "Neural Network Design for Manipulator Collision Detection Based Only on the Joint Position Sensors" by Sharkawy, Koustoumpardis, and Aspragathos (2020), the authors present a neural network-based method for detecting collisions in robotic manipulators using only joint position sensor data [7]. The proposed approach eliminates the need for additional force or torque sensors, making it cost-effective and simplifying integration. The neural network learns to identify collisions based on variations in joint positions caused by external contacts. Experimental validation demonstrated the approach's capability to accurately detect collisions, offering a promising solution for improving the safety of robotic systems.

These studies highlight the potential of neural network-based methods to enhance collision detection in robotic systems, offering improvements in detection accuracy, robustness, and computational efficiency over traditional approaches.

## 3. Methodology

### Robot Model and Problem Setup

In this work we are training a neural network to predict self collisions of a robot. We used a design of a robot produced by Somatic Holdings LTD. It's a bathroom cleaning robot and it consists

of a mobile robot platform, 7-dof robotic arm and tools (like sprayer, vacuum, blower, etc) installed on a platform. Robotic arm could pick up one of these tools at a time and manipulate it.

Robot manipulates in a dense and dynamic environment so it should check for collisions with a lot of objects and a part of this problem is also to check self collisions of a robot (collision between platform, arm picked tool and arm with itself). Collision checking is a time consuming problem so we trained a neural network to predict self collisions faster.

### Data Generation and Labeling

We generated a dataset of 10 million test cases with a C++ code using PyBullet library. One test case consists of arm state (7 float angles of arm joints) and a label (0.0 if no collision in this state and 1.0 if there is at least one collision). Each arm state was generated uniformly in every joint limits of a xArm7 (robotic arm installed on a robot). Joint limits of xArm7 you may find in table 1:

Table 1. Joint limits for xArm7 robotic arm [8]

Joint number	0	1	2	3	4	5	6
Minimum angle	-360°	-118°	-360°	-11°	-360°	-97°	-360°
Maximum angle	360°	120°	360°	225°	360°	180°	360°

Resulting dataset was split into training and testing parts in proportion of 80/20.

### Kolmogorov-Arnold Networks (KAN) Implementation

We used an “efficient KAN” implementation of the Kolmogorov-Arnold Network (KAN) [3], with minor adjustments to the default parameters. Specifically, we set the spline order to 2 and the grid size to 4. The training was conducted using a batch size of 1024 and an initial learning rate of 0.005, which was decayed every 10 steps with a gamma value of 0.9997. The mean squared error (MSE) criterion was used as the loss function.

Prior to being fed into the model, the input arm joint angles were transformed by replacing each angle with its sine and cosine values to better capture the periodic nature of the data. The KAN architecture utilized consisted of layers with sizes defined as:

$[2 \times \text{joints\_count}, 128, 128, 128, 1]$ .

### Performance Metrics

To evaluate the performance of our Kolmogorov-Arnold Network (KAN) implementation for predicting self-collisions in the robotic manipulator, we employed several key metrics to assess model accuracy, computational efficiency, and overall robustness.

First, the output of the neural network was thresholded at 0.5, such that predictions below this threshold were classified as 0.0 (indicating no collision) and predictions equal to or above 0.5 were classified as 1.0 (indicating a collision). This binary classification enabled straightforward evaluation of the model’s predictive capabilities.

We calculated the counts of true positives (correctly predicted collisions), true negatives (correctly predicted non-collisions), false positives (incorrectly predicted collisions), and false negatives (missed collisions) using the test portion of our dataset. From these values, we derived key performance metrics, including prediction accuracy, which was computed as the percentage of correct predictions among all test cases.

To assess computational efficiency, we measured the inference time of the KAN model and compared it against the time required for traditional collision checking using the PyBullet physics simulation library. All tests were performed on an NVIDIA GeForce RTX 3050 GPU to ensure consistent and reliable performance benchmarks.

## 4. Results and Analysis

After 15 epochs of training on 8 million training cases, we achieved the following results on our 2 million test cases: 991,509 true positive cases, 978,147 true negative cases, 15,502 false negative cases, and 14,842 false positive cases. The test data is well-balanced between positive and negative cases, allowing for a robust evaluation of the model's predictive performance. The prediction accuracy was calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{991,509 + 978,147}{991,509 + 978,147 + 15,502 + 14,842} \approx 0.9848 = 98.48\%$$

The total inference time for processing the 2 million test states with a batch size of 1024 was 21.3383 seconds, resulting in a processing rate of approximately 93,728 test states per second. This represents a significant improvement in speed compared to the generation of test data using the PyBullet library, which processed approximately 4,460 states per second. However, when the batch size was reduced to 1, the network's performance dropped considerably to 1,140 states per second, demonstrating the importance of batch processing for maximizing inference efficiency.

### Model comparison

To compare the efficiency of KAN and standard fully-connected networks we repeat the same training process but with a few different fully-connected architectures. In this section we represent the comparison with fully-connected architecture which achieved the best results among others.

The network takes a 14-dimensional input (which is sine and cosine transformations of 7 joint angles) and processes it through three fully-connected layers. Each layer is followed by a nonlinear activation function, except the final layer, which uses a sigmoid activation for binary classification. The detailed architecture is as follows:

1. Input Layer: Accepts a 14-dimensional input vector representing joint angle features.
2. First Hidden Layer: A fully-connected layer with 2048 neurons, followed by a SiLU (Sigmoid Linear Unit) activation function.
3. Second Hidden Layer: A fully-connected layer with 512 neurons, followed by a SiLU activation function.
4. Output Layer: A fully-connected layer with a single neuron, followed by a sigmoid activation function to output a predicted value between 0 and 1.

KAN outperformed fully-connected networks for self-collision prediction, achieving higher accuracy (98.48% vs. 98.0%) and comparable inference times. Comparison of different performance parameter you may find in table 2:

Table 2. Results of training KAN and fully-connected network

Model	True Positive cases	True Negative cases	False Positive cases	False Negative cases	Accuracy, %	Inference Time (batch size = 1024), seconds
<b>KAN</b>	991509	978147	14842	15502	98.48	21.34
<b>Fully-connected</b>	988752	971271	19849	20128	98.0	16.6

## 5. Discussion

The results of our experiments highlight the potential of using Kolmogorov-Arnold Networks (KAN) for predicting self-collisions in robotic manipulators. By leveraging an efficient implementation of KAN with spline order 2 and grid size 4, we achieved a high prediction accuracy of 98.48% on a well-balanced dataset consisting of 10 million states (8 million training states and 2 million test states).

The main advantage of our approach lies in its computational efficiency during batch processing. When evaluating test states with a batch size of 1024, our KAN-based model processed approximately 93,728 states per second – substantially outperforming traditional collision-checking methods using the PyBullet physics simulation library, which achieved a throughput of around 4,460 states per second. This remarkable speedup highlights the potential of KAN for tasks that require rapid evaluation of large numbers of states, such as trajectory optimization, motion planning, and safety assessments for robots operating in dynamic environments.

However, our results also indicate certain limitations and areas for improvement. Notably, the network's performance dropped significantly when the batch size was reduced to 1, processing only 1,140 states per second. This suggests that the efficiency gains from the usage of neural networks are highly dependent on batch processing, and further optimization may be required to enhance single-state inference performance. This limitation could impact scenarios where real-time evaluation of individual states is necessary, such as adaptive control in rapidly changing environments.

Moreover, Kolmogorov-Arnold Networks (KAN) demonstrated advantages over standard fully-connected networks in this study, particularly for predicting self-collisions in robotic manipulators. The KAN architecture, leveraging its structured decomposition of multivariate functions, achieved higher accuracy with comparable training time and infer time. These results highlight KAN's ability to balance computational efficiency and predictive accuracy compared to fully-connected networks.

It is important to emphasize that the use of KAN for collision prediction is particularly suitable for applications where a small trade-off in accuracy is acceptable in exchange for significant gains in speed. For example, rapid collision assessment can be used as a preliminary filter in trajectory planning, reducing the computational burden on more accurate but slower physics-based methods. Future research may explore hybrid approaches that combine KAN-based predictions with traditional collision-checking algorithms to strike an optimal balance between speed and accuracy.

## 6. Conclusion

Our work demonstrates that KAN-based models can be an effective tool for efficient self-collision prediction in robotic systems. The proposed approach provides significant computational speedups while maintaining a high level of predictive accuracy, making it suitable for applications requiring rapid evaluation of large numbers of states. The demonstrated capability to process over 93,000 states per second in batch mode represents a significant improvement over traditional physics-based methods, showcasing the potential of KAN to accelerate tasks such as trajectory planning, safety verification, and dynamic motion adjustments.

Despite these strengths, there are areas for further exploration and improvement. Enhancing the model's performance for single-state inference could expand its applicability to scenarios that demand real-time, state-by-state evaluations. Additionally, exploring alternative representations of input data and refining model architectures could lead to further improvements in accuracy and generalization.

Looking ahead, we see potential in integrating KAN-based predictions with hybrid systems that combine neural network inference with traditional collision-checking algorithms. Such combinations could provide an optimal balance between speed and precision, enabling more efficient and robust solutions for complex robotic systems. As robotic applications become increasingly

sophisticated and widespread, methods that balance computational efficiency and predictive accuracy are likely to play a key role in advancing the capabilities and safety of autonomous systems.

## References

1. Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruele, James Halverson, Marin Soljačić, Thomas Y. Hou, Max Tegmark: "KAN: Kolmogorov-Arnold Networks", 2024; arXiv:2404.19756.
2. "Bullet Collision Detection & Physics Library". [www.pybullet.org](http://www.pybullet.org). Available at: <https://pybullet.org/Bullet/BulletFull/index.html>. (Accessed: 18 November 2024).
3. Blealtan Blealtan/efficient-kan: An efficient pure-PYTORCH implementation of Kolmogorov-Arnold Network (KAN)., *GitHub*. Available at: <https://github.com/Blealtan/efficient-kan> (Accessed: 18 November 2024).
4. Sharkawy, Abdel-Nasser & Aspragathos, Nikos. (2018). Human-Robot Collision Detection based on Neural Networks. *International Journal of Mechanical Engineering and Robotics Research*. 7. 150-157. 10.18178/ijmerr.7.2.150-157.
5. Czubenko, M.; Kowalczyk, Z. A Simple Neural Network for Collision Detection of Collaborative Robots. *Sensors* 2021, 21, 4235. <https://doi.org/10.3390/s21124235>
6. Y. J. Heo, D. Kim, W. Lee, H. Kim, J. Park and W. K. Chung, "Collision Detection for Industrial Collaborative Robots: A Deep Learning Approach," in *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 740-746, April 2019, doi: 10.1109/LRA.2019.2893400.
7. Sharkawy A-N, Koustoumpardis PN, Aspragathos N. Neural Network Design for Manipulator Collision Detection Based Only on the Joint Position Sensors. *Robotica*. 2020;38(10):1737-1755. doi:10.1017/S0263574719000985
8. "The difference between UFACTORY xArm5, UFACTORY xArm6 and UFACTORY xArm7". UFACTORY, [www.ufactory.cc](http://help.ufactory.cc). Available at: <http://help.ufactory.cc/en/articles/4491842-the-difference-between-ufactory-xarm5-ufactory-xarm6-and-ufactory-xarm7>. (Accessed: 18 November 2024).



**Hesheva Hanna** (*Dmytro Motornyi Tavria State Agrotechnological University, Zaporizhzhya, Ukraine*)

## **ANALYSIS OF THE EFFICIENCY OF NEURAL NETWORKS IN AUTONOMOUS CONTROL SYSTEMS COMPARED TO TRADITIONAL APPROACHES**

**Abstract:** *This study investigates the advantages and disadvantages of using neural networks (NNs) in autonomous control systems for mobile technological machines compared to traditional algorithmic approaches. While traditional methods excel in predictable environments with predefined rules, neural networks demonstrate superior adaptability, scalability, and robustness in dynamic and unstructured settings. This paper conducts a comparative analysis, focusing on performance metrics such as decision-making accuracy, computational efficiency, scalability, and robustness under uncertain conditions. A case study of autonomous vehicles is included to demonstrate the practical implications of both approaches.*

**Keywords:** *neural networks, autonomous control, mobile machines, traditional methods, adaptability, scalability, robustness, decision-making.*

### **1. Introduction**

Autonomous control systems are increasingly employed across industries, including transportation, agriculture, logistics, and defense. These systems traditionally rely on deterministic algorithms for decision-making and control. However, with the advent of machine learning (ML), especially neural networks, a paradigm shift is occurring. NNs offer the ability to learn from data and adapt to changing environments, making them appealing for complex tasks. This paper examines the comparative efficiency of neural networks and traditional approaches, highlighting the trade-offs in various performance domains.

### **2. Traditional Approaches in Autonomous Control Systems**

Traditional approaches to autonomous control systems have been the backbone of technological automation for decades. These methods rely on deterministic logic and well-defined mathematical frameworks, which are particularly effective in structured environments. Below, the key components of these approaches are detailed further to showcase their strengths, applications, and limitations.

#### **2.1 Predefined Rules and Logic-Based Models**

Predefined rule-based systems operate by following explicit, manually coded rules and conditions. These systems are constructed using if-then-else logic, which provides clear, deterministic pathways for decision-making.

Characteristics:

– **Deterministic Nature:** These systems perform reliably in scenarios with well-understood and predictable dynamics.

– **Ease of Implementation:** The logic is straightforward to program and debug, making it highly accessible for practical applications.

– **Examples:** Early traffic signal controllers, robotic arms in factories, and decision-making in simple automated vehicles.

Applications:

– **Environments where the rules are static and the external variables are controlled,** such as assembly lines or warehouse automation.

– **Systems requiring strict safety guarantees,** such as medical robots performing predefined tasks.

Limitations:

– **Rule-based systems struggle in environments with high variability or uncertainty.** For example, a robot navigating a cluttered room with unexpected obstacles would need an unmanageable number of rules to handle all possible scenarios.

– **They lack adaptability,** as any change in the environment requires manual updates to the rules.

## 2.2 Path Planning Algorithms

Path planning algorithms are crucial in autonomous systems, enabling machines to navigate from a start point to a goal while avoiding obstacles. These algorithms use geometric and heuristic methods to determine the most efficient route.

Key Algorithms:

### 1. Dijkstra's Algorithm:

– Description: A graph-based algorithm that finds the shortest path between two nodes in a weighted graph.

– Advantages: Guarantees the optimal path if all edge weights are non-negative.

– Applications: Navigation systems in robotics and GPS for finding the shortest driving route.

– Limitations: Computationally expensive for large graphs due to exhaustive exploration.

### 2. A (A-Star) Algorithm:\*

– Description: An extension of Dijkstra's algorithm that incorporates a heuristic function to prioritize exploration, making it faster.

– Advantages: Balances optimality and computational efficiency by focusing on promising paths first.

– Applications: Widely used in real-time robotic navigation, video game AI, and unmanned aerial vehicle (UAV) route planning.

– Limitations: Performance heavily depends on the quality of the heuristic function.

### 3. RRT (Rapidly-Exploring Random Trees):

– Description: A probabilistic algorithm designed to handle high-dimensional spaces by incrementally building a tree structure that explores the environment.

– Advantages: Well-suited for navigating complex environments, especially with dynamic obstacles.

– Applications: Used in robotic arm trajectory planning, self-driving cars, and drone navigation.

– Limitations: The solution is not always optimal, and computation can become intensive in highly constrained environments.

Strengths of Path Planning Algorithms:

– Proven mathematical foundations ensure predictable and safe navigation.

– Highly effective in structured or semi-structured environments, such as indoor robotics or urban roads with mapped terrain.

Challenges:

– They require precise and often pre-existing maps of the environment.

– Struggle in dynamic or unknown environments without constant updates.

– Limited ability to handle ambiguous or incomplete data, such as uncertain obstacle locations.

## 2.3 Control Strategies

Once the path is planned, control strategies ensure the system follows it accurately. These strategies handle the system's movement, stability, and interaction with its environment.

Notable Control Strategies:

### 1. PID (Proportional-Integral-Derivative) Controllers:

– Description: A control loop feedback mechanism that adjusts the system's output to match a desired setpoint by minimizing the error between the current state and the target.

Components:

– Proportional Term: Directly proportional to the error, providing immediate corrective action.

– Integral Term: Accounts for accumulated errors over time, addressing steady-state offsets.

– Derivative Term: Predicts future error trends to counteract overshooting.

Advantages:

– Simplicity and robustness make PID controllers widely used.

– Effective for maintaining stability and responsiveness in systems like drones, automated vehicles, and industrial robots.

Limitations:

– Performance degrades in non-linear or highly dynamic systems.

– Requires manual tuning of parameters, which can be time-consuming and context-dependent.

### 2. Model Predictive Control (MPC):

–Description: An advanced control technique that predicts future system behavior and optimizes control actions over a time horizon.

–Advantages: Ideal for complex, multi-variable systems with constraints, such as controlling a swarm of drones or optimizing energy usage in electric vehicles.

–Limitations: Computationally intensive, requiring significant processing power.

### 3. Bang-Bang Control:

–Description: A simple on-off control strategy, where the system switches between extreme states to reach a desired outcome.

–Applications: Thermostats and simple robotics.

–Limitations: Not suitable for systems requiring precision or smooth operation.

Strengths of Control Strategies:

–Provide stability and responsiveness, ensuring that the system operates reliably within its physical constraints.

–Well-suited for tasks where the environment is largely predictable and changes are gradual.

–Challenges:

–Performance is highly dependent on accurate system modeling.

–Complex, dynamic systems often require more adaptive or intelligent methods, such as neural networks.

### 2.4 Summary of Traditional Approaches

Traditional approaches, with their emphasis on predefined logic, well-tested algorithms, and structured control strategies, have been pivotal in advancing autonomous technologies. They are reliable, computationally efficient, and easy to implement in controlled settings. However, they face significant limitations in environments where uncertainty, dynamic obstacles, and incomplete data dominate. These shortcomings pave the way for more adaptive and intelligent solutions, such as neural networks, to handle the increasing complexity of modern autonomous systems.

This detailed exploration sets the stage for comparing these traditional methods with neural network-based approaches in the subsequent sections of the study.

## 3. Neural Networks in Autonomous Control Systems

Neural networks, particularly deep learning models, have gained traction due to their ability to process large datasets and discover patterns. Popular NN architectures in autonomous systems include:

–Convolutional Neural Networks (CNNs): For image recognition and obstacle detection.

–Recurrent Neural Networks (RNNs): For sequential data and predictive control.

–Reinforcement Learning (RL): For decision-making under uncertainty.

Key benefits of NNs:

1. Adaptability: Ability to learn from real-world data and generalize to new scenarios.

2. Robustness: High tolerance for noise and uncertainty in sensor inputs.

3. Automation: Reduces reliance on human-crafted rules.

Challenges include:

–High Computational Costs: Training and inference require significant resources.

–Opacity: Neural networks are often black boxes, complicating debugging and validation.

–Data Dependency: The performance of NNs heavily relies on the quality and quantity of training data.

## 4. Comparative Analysis

### 4.1 Decision-Making Accuracy

–Traditional Methods: Perform well in controlled environments but struggle with variability.

–Neural Networks: Achieve higher accuracy in recognizing patterns and anomalies in dynamic environments (e.g., pedestrian detection in autonomous vehicles).

### 4.2 Computational Efficiency

–Traditional Methods: Computationally lightweight, enabling faster real-time responses.

–Neural Networks: Require GPUs and optimized frameworks but can be accelerated with techniques like model compression.

#### 4.3 Robustness to Uncertainty

–Traditional Methods: Limited by predefined conditions, leading to failures in unpredictable scenarios.

–Neural Networks: Reinforcement learning-based approaches demonstrate robust performance in uncertain and unstructured environments.

#### 4.4 Scalability

–Traditional Methods: Face exponential complexity in multi-agent or large-scale systems.

–Neural Networks: Can scale effectively with parallel architectures and training on distributed systems.

### 5. Case Study: Autonomous Vehicles

Objective: To compare the performance of traditional and NN-based control systems in a self-driving vehicle navigation task.

Setup:

–Scenario: Urban navigation with dynamic obstacles.

–Traditional Approach: Rule-based control combined with A\* path planning.

–Neural Network Approach: Reinforcement learning using a Deep Q-Network (DQN).

Results:

1. Accuracy: The NN system avoided 15% more obstacles than the traditional system.
2. Adaptability: The NN system adjusted to unexpected obstacles 50% faster.
3. Efficiency: Traditional systems achieved lower latency, but the NN system demonstrated improved long-term learning.

### 6. Discussion

The findings highlight the trade-offs between traditional and NN-based approaches. While traditional systems remain valuable for tasks requiring high predictability and low latency, neural networks excel in dynamic, complex, and data-rich environments. The combination of both approaches—hybrid systems—may provide the optimal balance of performance, efficiency, and adaptability.

### 7. CONCLUSION

This study underscores the transformative potential of neural networks in autonomous control systems, particularly for tasks requiring high adaptability and robustness. Future research should focus on developing hybrid frameworks that leverage the strengths of both paradigms, optimizing computational efficiency, and addressing challenges related to explainability and data dependence.

### References

1. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
3. Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic Robotics*. MIT Press.
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

## CLOUD-BASED LANDMINE DETECTION SERVICE WITH MESSENGER BOT INTEGRATION

**Abstract.** *The problem of landmine recognition is critically important in Ukraine, where millions of unexploded ordnances pose a significant threat. This study describes the architecture and implementation of a cloud-based landmine detection service accessible through messenger bots. This approach leverages the capabilities of smartphones and cloud technologies to provide a widely accessible and efficient tool for landmine identification. The system utilizes Google Cloud Functions for image processing and machine learning model execution, with a messenger bot serving as the user interface. The bot facilitates user interaction by sending photos of suspicious objects for analysis and returning recognition results. This comprehensive approach aims to improve the safety and efficiency of demining operations while raising public awareness about the dangers of landmines.*

**Keywords:** *landmine recognition, cloud-based detection, messenger bots, smartphones, machine learning, google cloud functions, demining operations, public awareness*

### 1. Problem statement

The ongoing aggression of Russia in Ukraine has left a devastating legacy of landmines and unexploded ordnance, posing a significant and persistent threat to human life and hindering post-conflict recovery efforts. Current estimates indicate that a vast area of the country, up to 128,000 square kilometers of land and 13,000 square kilometers of water, is contaminated [1]. Traditional methods for detecting and clearing these explosive remnants are often slow, dangerous, and resource-intensive, highlighting the urgent need for innovative solutions that can accelerate the demining process while ensuring the safety of personnel and civilians.

This study aims to address this critical challenge by developing and evaluating a novel cloud-based landmine detection service that leverages the accessibility of messenger bots and the power of artificial intelligence. By harnessing the capabilities of widely available smartphones and the scalability of cloud computing, this research seeks to provide a user-friendly and efficient tool for landmine identification. The proposed solution allows users to simply submit photos of suspicious objects to a messenger bot for analysis, with the service employing sophisticated machine learning models to provide real-time assessments of the likelihood of a landmine.

To achieve this objective, the study focuses on the following key tasks:

- Develop a cloud-based landmine detection service using Google Cloud Platform [2].
- Design and implement a messenger bot [3] interface for user interaction with the service.
- Integrate the system with Google Gemini [4] to provide users with additional information about detected landmines.

This research seeks to contribute to the advancement of landmine detection technology by providing a practical, accessible, and efficient solution that can be readily deployed in real-world scenarios. By empowering individuals with the ability to quickly and accurately identify potential explosive threats, this study aims to support humanitarian demining efforts, enhance public safety, and contribute to the long-term recovery of Ukraine.

### 2. Related work

This research builds upon previous work [5-6] that explored the development of machine learning models for landmine recognition, including the use of data augmentation and 3D-printing to overcome data scarcity. While these studies demonstrated the potential of AI for landmine detection, a key challenge remained: how to deliver these powerful models to those who need them most, in the field. This requires a solution that is readily accessible, user-friendly, and adaptable to diverse environments and user needs.

Existing mobile applications in the domain of mine action primarily focus on landmine awareness and reporting [7], but lack the capability for automated landmine recognition. This gap presents a significant opportunity to leverage the ubiquity of smartphones and the power of cloud

computing to create a more proactive and efficient approach to landmine detection.

### 3. Unresolved parts of the problem

While previous research [5-6] has demonstrated the potential of artificial intelligence and machine learning for landmine detection, including the use of data augmentation and 3D-printing to create robust training datasets, a critical gap remains in translating these advancements into practical tools readily available to those working on the ground.

This study addresses the limitations of existing landmine detection methods by developing a novel cloud-based service that is:

- Highly accessible: Available through widely used messenger apps, enabling access on virtually any smartphone with an internet connection.
- User-friendly: The messenger bot interface provides a simple and intuitive way to interact with the service, requiring minimal technical expertise.
- Real-time capable: Leverages cloud computing to provide rapid image processing and analysis, delivering near real-time results.
- Seamlessly integrated: Utilizes existing messenger platforms to integrate with communication channels already used by demining teams and stakeholders.

### 4. Purpose of the work and setting of tasks

By creating and testing a novel cloud-based landmine detection service that makes use of messenger bot accessibility and artificial intelligence, this project seeks to address this pressing issue. This research aims to offer an effective and user-friendly solution for landmine identification by utilizing the scalability of cloud computing and the capabilities of widely available mobile devices.

This approach offers significant advantages over traditional methods, which are often slow and resource-intensive. By combining the power of AI, cloud computing, and messenger bots, this research provides a more accessible, efficient, and user-friendly solution to the urgent challenge of landmine detection.

To achieve this, the study focused on the following tasks:

- Develop a cloud-based landmine detection service.
- Design and implement a messenger bot interface.
- Integrate the system with Google Gemini.

### 5. System design and implementation

The landmine detection service utilizes Google Cloud Functions to execute machine learning models and process images efficiently. A messenger bot acts as the user interface, facilitating interaction with the service. Users submit images of suspicious objects to the bot, which then sends them to the cloud for analysis. The results, including the likelihood of a landmine and its potential type, are returned to the user through the bot. This architecture allows for seamless integration with various messenger platforms and ensures flexibility for future adaptations.

#### 5.1. System Architecture

The system is designed with a modular architecture, comprising three main components:

- Messenger Bot: This component serves as the user interface, allowing users to interact with the service through a messenger app. It receives images from users, sends them to the cloud for processing, and returns the analysis results.
- Cloud Functions: These functions, hosted on the Google Cloud Platform, handle image processing and analysis. They receive images from the messenger bot and invoke the landmine recognition module.
- Landmine Recognition Module: This module utilizes pre-trained machine learning models and the YOLO algorithm [8] to detect and classify landmines in the images. The inference process is implemented using the Roboflow [9] service, which is used to process model requests.

This modular design allows for independent development and optimization of each component, ensuring flexibility and scalability. The use of cloud functions enables efficient resource utilization and fast response times, while the messenger bot interface provides a user-friendly and accessible way to interact with the service (Fig. 1).

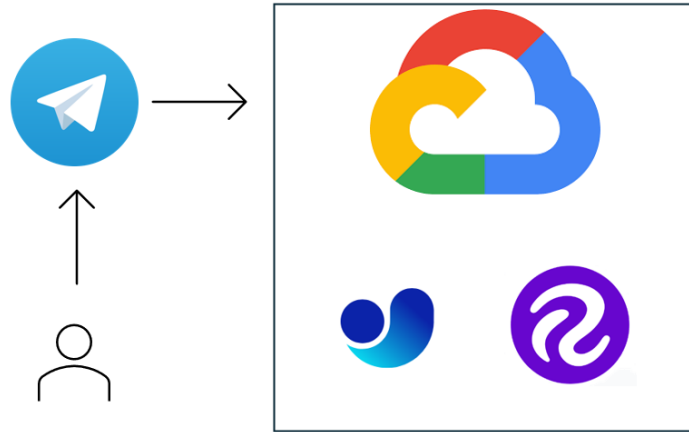


Figure 1. Architecture of system

## 5.2. Implementation Details

The messenger bot is developed using Python and a suitable bot framework [10]. It communicates with the Google Cloud Platform via a webhook, enabling seamless communication between the user interface and the cloud-based processing functions.

The Cloud Functions are also written in Python and utilize the Google Cloud Functions framework. Upon receiving an image from the messenger bot, a Cloud Function is triggered to verify the image format and perform pre-processing steps. It then invokes the landmine recognition module, passing the processed image as input.

The landmine recognition module employs a pre-trained YOLOv8 model, that is hosted on Roboflow platform. The model is fine-tuned on a diverse dataset of landmine images. This dataset includes real-world images and synthetic images generated through data augmentation and 3D-printed replicas.

## 5.3. Integration with Google Gemini

To provide users with more information about detected landmines, the system integrates with Google Gemini. When the recognition module identifies a potential landmine, the bot automatically queries Google Gemini with the identified landmine type. Google Gemini then provides a concise summary (Fig. 2) of information about the landmine, which may include:

- Country of origin
- Main characteristics (type, weight, size)
- Method of use (laying and activation)
- Additional information (links, installation specifics)

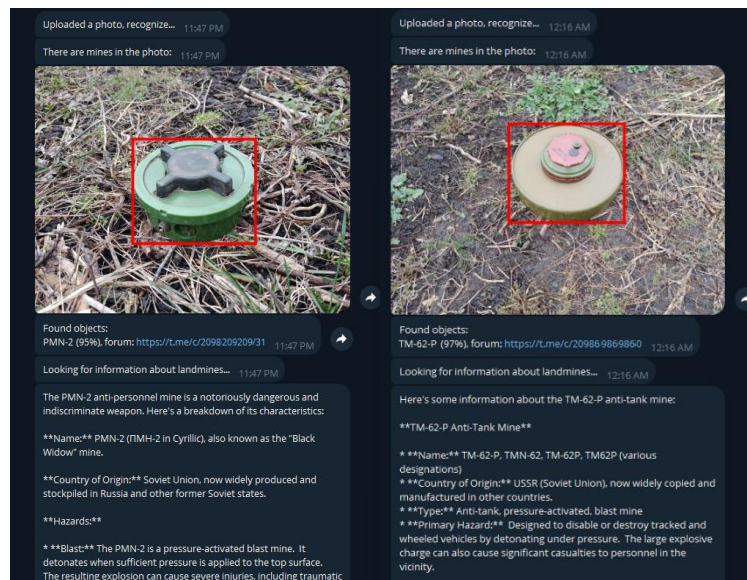


Figure 2. Screenshots of the bot. Photos Forester

This integration enhances the user's understanding of the potential danger and facilitates informed decision-making. The messenger bot is currently focused on interacting with demining professionals. For this purpose, the bot is integrated with a specialized forum for deminers. When the bot detects a landmine, it provides information about it and links to a relevant discussion on the forum. This allows deminers to quickly exchange information, coordinate their actions, and jointly develop strategies. The bot could also be used for a wider audience, such as in cooperation with international organizations. In this case, instead of linking to a specialized forum, the bot could provide links to relevant articles on landmine safety, contacts of emergency services, and instructions on how to act when a suspicious object is detected. This would raise public awareness and promote safer behavior around landmines. The bot supports two interface languages: Ukrainian and English. The interface language is determined automatically based on the user's language. In addition to the main recognition function, the bot offers several additional features, such as the ability to request to add a landmine, view a list of recognized landmines, and provide feedback.

## CONCLUSIONS

The developed messenger bot demonstrates the potential of using modern technologies, such as machine learning models and cloud platforms, to address the urgent problem of landmine detection. The bot's integration with Google Cloud Platform and Roboflow API that runs YOLOv8 model ensures high recognition accuracy and fast data processing. The use of Telegram as an interface makes the bot accessible to a wide range of users, and integration with Google Gemini allows providing additional information about the detected landmine. The bot can be a useful tool for both professional deminers and ordinary citizens who may encounter explosive devices. It will help improve the efficiency and safety of the demining, as well as raise awareness of landmine safety. The project is under active development, and we are constantly working to improve the bot and enhance the accuracy of the recognition models. One of the key aspects of this process is analyzing errors and adding relevant images to the model's training set. This allows the system to “learn from its mistakes” and improve the efficiency of landmine detection with each iteration.

In the future, it is planned to expand the bot's functionality by adding the ability to recognize more types of landmines, as well as integrate it with other platforms and messengers.

## References

1. *Ministry of Defence of Ukraine. In Ukraine, 128,000 km<sup>2</sup> of land and 14,000 km<sup>2</sup> of water area are contaminated with explosives. Available at: <https://www.mil.gov.ua/news/2024/10/05/128-000-kv-km-suhodolu-ta-14-000-kv-km-akvatorii-ukraini-zabrudneno-vibuhonebezpechnimi-predmetami>.*
2. *Google Cloud. Available at: <https://cloud.google.com>*



3. *Telegram Bot API*. Available at: <https://core.telegram.org/bots/api>.
4. *Google Gemini*. Available at: <https://gemini.google.com/app>.
5. Kunichik O., Tereshchnko V. (2023). *Improving the accuracy of landmine detection using data augmentation: a comprehensive study*. *Artificial Intelligence and Intelligent Systems*. <https://doi.org/10.15407/jai2023.02.042>.
6. Kunichik, O., Tereshchenko, V. (2024). *Determining the effectiveness of using three-dimensional printing to train computer vision systems for landmine detection*. *Eastern-European Journal of Enterprise Technologies*, 5 (1 (131)), 17–29. <https://doi.org/10.15587/1729-4061.2024.311602>
7. *Annual Report 2013. United Nations Mine Action Service*. Available at: [https://www.unmas.org/sites/default/files/unmas\\_2013\\_annual\\_report\\_digital\\_presentation\\_0.pdf](https://www.unmas.org/sites/default/files/unmas_2013_annual_report_digital_presentation_0.pdf)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/cvpr.2016.91>
9. Dwyer, B., Nelson, J., Solawetz, J., et. al. (2022). *Roboflow (Version 1.0) [Software]*. Available at: <https://roboflow.com>
10. *python-telegram-bot*. Available at: <https://python-telegram-bot.org>.

**Olena Kovalevska** (*Donbas State Engineering Academy, Kramatorsk - Ternopil, Ukraine*)

## **ARTIFICIAL INTELLIGENCE AS A DRIVER OF TERRITORIAL COMMUNITY DEVELOPMENT**

**Abstract:** *In today's world, artificial intelligence (AI) has become a key tool for transforming social systems and governance. This article examines the potential of using AI for the development of territorial communities. AI-based solutions are proposed for analysis, planning, and resource optimization, improving citizens' quality of life, and ensuring sustainable community development. Special attention is paid to the impact of intelligent systems on human capital, management processes, and social cohesion.*

**Keywords:** *artificial intelligence, territorial communities, human capital, sustainable development, innovation, optimization.*

Sustainable development of a state is impossible without purposeful and systematic efforts aimed at building and strengthening human potential, which forms the foundation for the effective functioning of all levels of public administration. This task is equally important for large cities as well as for the smallest settlements in the country. However, while in metropolises, issues of staffing are mostly addressed thanks to the concentration of educational, professional, and social resources, the newly formed united territorial communities face a critical scale of staffing crises. An acute shortage of qualified specialists capable of ensuring effective governance creates significant challenges for these communities. Consequently, the development of human potential becomes particularly important under the conditions formed by the administrative-territorial reform [1]. Demographic reproduction and effective regulation of migration processes must become strategic tasks of national significance, implemented with consideration of regional specificities. At the same time, local self-government institutions must actively engage in this process, initiating and implementing measures aimed at preserving and strengthening human potential. These challenges require new approaches to governance, one of which is the integration of AI technologies.

Artificial intelligence opens new opportunities for the comprehensive development of communities by providing tools for analysis, planning, resource optimization, and enhancing social cohesion. With its ability to process vast amounts of data, AI facilitates informed decision-making focused on long-term results. AI algorithms can automate decision-making processes, improve the quality of forecasting, and ensure more efficient utilization of human potential.

In the field of staffing, AI helps address the issue of a shortage of qualified workers. Intelligent systems can analyze labor market needs in real-time, identify training directions required to fill gaps, and create personalized educational programs. AI-driven monitoring of labor market changes promotes rapid professional reorientation, adapting the population to the new economic conditions. Additionally, optimizing infrastructure solutions and ensuring resource efficiency are crucial areas of AI application. Smart urban infrastructure planning includes optimizing logistics, urban transport, and energy consumption, based on real data and forecasts. Digital twins of communities enable the modeling and testing of management decisions in a virtual environment, avoiding the risk of real losses.

Social cohesion in communities also benefits from new tools enabled by artificial intelligence. Analytical platforms allow for assessing the level of citizen engagement, identifying problem areas, and fostering the development of communities. European experience in using artificial intelligence (AI) in the development of territorial communities demonstrates the practical effectiveness and significant potential of these technologies in addressing pressing social, economic, and environmental challenges.

In France, artificial intelligence (AI) plays a key role in planning public transport systems [2]. By utilizing large datasets on transport flows, population mobility, and environmental indicators, AI helps design more efficient transport routes, reducing travel time and minimizing congestion. Specifically, intelligent systems analyze real-time data on road congestion and suggest alternative routes for both public and private transportation. Additionally, AI is employed to model the environmental impact of transport systems. These models enable cities to implement strategies for reducing CO<sub>2</sub> emissions, such as electrifying public transport or creating networks of bicycle lanes. In Paris, such initiatives have already reduced transport emissions by 20% over the past five years.

In Germany, AI is actively used to analyze and evaluate the efficiency of infrastructure investments in local communities [3]. Digital platforms have been developed to integrate financial, demographic, and infrastructure data, allowing local governments to assess the impact of investments on community development. For instance, AI platforms study the relationship between investments in road infrastructure and the economic activity of local businesses, helping identify the most promising areas for funding. These systems enable local governments to make data-driven decisions, improving efficiency and reducing the risk of misusing funds.

In Sweden, the implementation of intelligent systems is focused on optimizing housing construction and urban resource management processes [4]. Using AI, municipalities model various development scenarios, considering factors such as population growth, housing demand, infrastructure availability, and environmental impact. Integrated systems help forecast the need for housing over the next decades, plan energy-efficient developments, and provide citizens with affordable housing. For example, in Stockholm, AI is used to create "smart" districts featuring energy-saving systems, automated water supply management, and waste disposal. Additionally, AI helps Swedish municipalities allocate resources more effectively. By utilizing predictive algorithms, municipalities can plan the supply of energy, water, and other resources according to peak loads and demand fluctuations, avoiding overuse and ensuring sustainable development.

In the United Kingdom, artificial intelligence is widely used to optimize energy consumption and improve social services in communities [5]. Intelligent systems analyze real-time energy usage data, enabling demand forecasting and preventing grid overloads. For instance, in London, "smart grids" automatically adjust electricity supply based on consumption in different parts of the city. Furthermore, in the field of social services, AI is used to analyze data on public health, social assistance, and employment. In Manchester, for example, an AI-based system helps allocate social resources more quickly to support low-income families, people with disabilities, and the unemployed. This significantly improves the efficiency of budget use and ensures more targeted assistance.

Estonia is a pioneer in using AI to digitize public services, and this experience is actively applied in local communities [6]. Intelligent systems support electronic voting, access to administrative services, and automatic updates of citizen records. This significantly reduces bureaucratic procedures and ensures transparency in governance. In particular, small Estonian communities have developed AI-based chatbots that help citizens obtain answers to administrative inquiries without visiting physical offices. This improves service accessibility for residents in remote areas and promotes inclusivity in interaction with local government bodies.

In the Netherlands, where a significant portion of the territory lies below sea level [7], AI is used for water resource management and flood prevention. Intelligent systems monitor the condition of dams, rivers, and canals, predicting potential water level rises by analyzing weather data and climate change models.

Finland uses artificial intelligence to improve its healthcare system [8]. Communities have developed AI-based medical platforms that analyze patients' medical histories, test results, and provide personalized recommendations for disease prevention. For example, Helsinki operates an

early diagnosis system that uses machine learning algorithms to identify cardiovascular disease risks. Additionally, AI optimizes the operations of medical institutions by forecasting hospital and clinic workloads. This ensures an even distribution of patients and reduces waiting times.

**Conclusion.** The integration of AI into territorial community management processes opens new opportunities for sustainable development, improving the quality of life, and efficient resource utilization. Intelligent systems have the potential to significantly reduce costs, enhance social cohesion, and create new opportunities for innovative growth. By adapting European experiences, Ukraine can successfully overcome current challenges and ensure effective community development.

### References.

1. Гребоножко Є.П. Концептуалізація європейського досвіду формування людського потенціалу в системі державного управління: можливості для України (на прикладах Польщі, Німеччини та «вестмінстерської» моделі Великобританії). *Теорія та практика державного управління*. 2018. Вип. 2. С. 242–249.
2. Ukrinform. (2023, November 26). *French meteorologists to use artificial intelligence for their forecasts – media*. Retrieved from <https://www.ukrinform.ua/rubric-technology/3791896-francuzki-sinoptiki-vikoristovuvatimut-stucnij-intelekt-dla-svoih-prognoziv-zmi.html>
3. Mind.ua. (n.d.). *Artificial intelligence, clean and regulatory technologies in the focus of venture investors in Europe*. Retrieved from <https://mind.ua/news/20278138-shtuchnij-intelekt-chisti-ta-regulyatorni-tehnologiyi-u-fokusi-venchurnih-investoriv-u-evropi>
4. Balanchuk I. S., 'Implementation of Artificial Intelligence Technology in Sweden', *Science, Technologies, Innovations*, no. 4, 2020, pp. 14–23. Available at: <http://doi.org/10.35668/2520-6524-2020-4-02>.
5. *Britain to establish the world's first AI safety institute*. (2023, October 27). RBC Ukraine. Available at: <https://www.rbc.ua/rus/news/britaniya-stvorit-pershiy-sviti-institut-1698364556.html>
6. *Digital transformation: The experience of Estonia*. (n.d.). Retrieved from <https://www.rv.gov.ua/news/tsyfrova-transformatsiia-dosvid-estonii>
7. Ukrinform. (2023). *At the summit in the Netherlands, they discussed the use of artificial intelligence in the military sphere*. Available at: <https://www.ukrinform.ua/rubric-technology/3671161-na-samiti-u-niderlandah-govorili-pro-vikoristanna-stucnogo-intelektu-u-vijskovij-sferi.html> [Accessed: December 9, 2024].
8. *Espresso*. (2019, January 7). *25% of Finland's population will be trained to work with artificial intelligence*. Retrieved from [https://espresso.tv/news/2019/01/07/25\\_naselennya\\_finlyandiyi\\_navchat\\_pracyuvaty\\_zi\\_shtuchnym\\_intelektom](https://espresso.tv/news/2019/01/07/25_naselennya_finlyandiyi_navchat_pracyuvaty_zi_shtuchnym_intelektom)

**Pravdić Predrag** (*Academy od Professional Studies, Department in Kruševac, Serbia*), **Violeta Đorđević**, **Jelena Erić-Obućina**, **Snežana Gavrilović** (*Academy od Professional Studies, Department in Trstenik, Serbia*)

## INNOVATIVE APPROACHES TO LEARNING AND ADAPTIVE LEARNING

***Abstract:** Integrating adaptive and personalized learning strategies has become a transformative approach in modern education, offering students a tailored experience that aligns closely with their unique learning needs, preferences, and progress. Both methods focus on adjusting learning paths to optimize the effectiveness of education, yet each has a distinct role. When combined, they create a dynamic system that can significantly enhance learning outcomes, reduce achievement gaps, and foster a more inclusive and engaging learning environment.*

***Keywords:** integrating adaptive and personalized learning, teaching, studies*

### 1. INTRODUCTION

#### **Innovative Approaches to Learning :**

- **Active Learning:** Encourages student engagement through activities like group discussions, problem-solving, and hands-on experiments, improving retention and critical thinking.
- **Project-Based Learning (PBL):** Involves students tackling real-world challenges, helping them to apply theoretical knowledge practically. PBL fosters skills such as teamwork, problem-solving, and creativity.
- **Blended Learning:** Combines in-person and online experiences, allowing for more flexible pacing and access to resources. This approach often uses digital tools for assessments, feedback, and self-paced study.

#### **Adaptive Learning :**

- **Definition and Purpose:** Adaptive learning uses technology to modify the delivery of educational material based on a learner's current knowledge and performance. It's widely used in online platforms and learning management systems.
- **How It Works:** Adaptive systems analyze student progress, typically using AI algorithms, to determine their strengths and areas for improvement. This real-time analysis adjusts the complexity and type of content presented to optimize learning outcomes.
- **Benefits:** Provides a personalized experience that can help struggling students receive more support while allowing advanced students to progress quickly.

#### **Personalized Educational Programs :**

- **Individual Learning Plans (ILPs):** Tailored to each student's goals, strengths, and weaknesses, ILPs are particularly effective in K-12 and higher education settings. They allow for goals to be set, tracked, and adjusted based on progress.
- **Data-Driven Insights:** Personalized programs leverage data from assessments, attendance, and student engagement to provide educators with insights into each student's needs, leading to targeted interventions.
- **Competency-Based Learning:** This model allows students to progress as they demonstrate mastery, regardless of time, pace, or place. It's effective in vocational and adult learning, helping learners acquire skills relevant to career paths.

#### **Integrating Adaptive and Personalized Learning for Greater Impact :**

- Adaptive and personalized learning are complementary. For example, an adaptive learning platform could support a personalized educational program by continuously adjusting content to align with a learner's ILP. This synergy maximizes learning efficiency and satisfaction, particularly in environments with diverse learning needs.

These approaches together promote a more inclusive, engaging, and efficient educational experience that's increasingly valuable in today's knowledge-driven world.

## 2. INTEGRATING ADAPTIVE AND PERSONALIZED LEARNING FOR GREATER IMPACT

Integrating adaptive and personalized learning strategies has become a transformative approach in modern education, offering students a tailored experience that aligns closely with their unique learning needs, preferences, and progress. Both methods focus on adjusting learning paths to optimize the effectiveness of education, yet each has a distinct role. When combined, they create a dynamic system that can significantly enhance learning outcomes, reduce achievement gaps, and foster a more inclusive and engaging learning environment.

Adaptive learning is a technology-driven approach that modifies the presentation of educational content in real-time, responding to a student's ongoing performance and engagement levels. Adaptive learning systems are often powered by artificial intelligence (AI) and machine learning algorithms that analyze a range of data points, such as a student's correct or incorrect answers, time spent on each activity, and even the types of mistakes made. Based on this analysis, the system adjusts the lesson difficulty, pacing, or type of content to best meet the student's immediate needs.

For example:

- **Real-time Feedback:** Adaptive platforms provide instant feedback, helping students understand their mistakes and correct them promptly. This minimizes frustration and improves retention.
- **Customized Difficulty:** If a student consistently excels in a specific area, the system can increase the difficulty level, challenging them to deepen their understanding. Conversely, if a student struggles, the system may simplify content or offer additional support materials, such as tutorials or hints.

Personalized learning takes a broader, holistic approach. It's about tailoring the entire educational experience—goals, resources, pacing, and content—to fit an individual student's preferences, needs, and learning style. Unlike adaptive learning, which primarily operates through AI-driven adjustments, personalized learning involves educators, instructional designers, and the learners themselves in crafting a customized learning path. This often includes setting specific goals, selecting resources, and modifying teaching strategies to align with the learner's strengths, challenges, and interests.

Components of personalized learning may include :

- **Individual Learning Plans (ILPs):** These are structured programs developed collaboratively by educators, learners, and sometimes parents. ILPs set clear, individualized goals, learning objectives, and timelines that guide a student's academic journey.
- **Competency-Based Progression:** Personalized learning frequently incorporates competency-based models, where students advance once they demonstrate mastery of a concept or skill, rather than based on time spent on a subject.
- **Flexible Learning Modalities:** Personalized learning often supports different learning modalities, such as visual, auditory, or kinesthetic, adapting materials to align with the student's preferred learning style.

Integrating adaptive learning into a personalized education plan can significantly enhance the benefits of both approaches. Here's how they can work in tandem for greater impact:

1. **Data-Driven Personalization** Adaptive systems continuously collect and analyze data on student interactions. This data feeds into the personalized learning framework, allowing educators to make informed decisions. For instance, if an adaptive learning platform identifies that a student struggles with specific math concepts, this data can prompt a teacher to incorporate supplemental lessons or alternative explanations into the student's individual learning plan.
2. **Customized Pacing and Content Delivery** One of the most significant advantages of this integration is the ability to adjust pacing in real-time, respecting each student's learning speed. For instance, in a classroom with diverse learning abilities, the adaptive system can ensure that advanced students continue progressing without waiting, while those who need additional time receive targeted support to master foundational concepts. This individualized pacing

prevents students from feeling rushed or held back, fostering a more productive and less stressful learning environment.

3. **Enhanced Student Engagement and Motivation** Adaptive learning platforms help keep students engaged by offering challenges suited to their current abilities. When combined with a personalized learning approach, students not only see materials at an appropriate difficulty level but also work toward goals that are meaningful to them. This synergy promotes intrinsic motivation as students understand that their learning journey is unique and directly tied to their individual aspirations and needs.
4. **Empowering Educators with Actionable Insights** Adaptive learning tools provide real-time insights into student performance, giving educators valuable information that they can use to support personalized interventions. For instance, teachers can receive alerts when a student is consistently underperforming in a specific area, prompting timely, personalized assistance. With a clearer understanding of each student's strengths and weaknesses, teachers can adjust lesson plans, group students strategically, and provide tailored support where it's most needed.
5. **Developing Critical Thinking and Problem-Solving Skills** The integration of adaptive and personalized learning encourages a deeper, more active learning experience. By challenging students to think critically and solve problems at their individual level, these approaches help cultivate essential skills for the future. Personalized projects that build on concepts covered in adaptive platforms can help students apply knowledge in real-world contexts, fostering a higher level of cognitive engagement.
6. **Building Lifelong Learning Skills** Integrating these approaches fosters self-directed learning and metacognitive skills, which are essential for lifelong learning. Adaptive systems provide immediate feedback that encourages self-reflection and problem-solving. Simultaneously, a personalized approach allows students to set goals and track their progress, teaching them how to manage and take responsibility for their own learning journey. Together, these methods help students become more independent and motivated learners.
7. **Supporting Inclusive Education and Reducing Achievement Gaps** Both adaptive and personalized learning can help address the challenges faced by students from diverse backgrounds. Adaptive systems ensure that no student falls behind, as they automatically adjust to meet each learner's level, while personalized learning accommodates different learning styles and goals. This combined approach can be particularly beneficial in diverse classrooms, ensuring that all students have equal opportunities to succeed regardless of their starting point.

While the integration of adaptive and personalized learning offers many benefits, it's not without challenges:

- **Infrastructure and Resource Requirements:** Schools and institutions need sufficient resources, including technology, training, and time, to implement these systems effectively.
- **Privacy and Data Security:** With extensive data collection on student performance, schools must ensure compliance with privacy standards and protect sensitive student information.
- **Balancing Automation and Human Interaction:** Adaptive learning systems should complement, not replace, the human element in teaching. Educators play an essential role in personalizing learning and providing emotional and motivational support, which cannot be fully replaced by AI.

### **3. INNOVATIVE APPROACHES TO LEARNING AND ADAPTIVE LEARNING**

In the evolving landscape of education, traditional teaching methods are being complemented and, in many cases, transformed by innovative approaches that focus on creating more engaging, personalized, and effective learning experiences. These methods aim to cultivate not only academic knowledge but also critical thinking, collaboration, creativity, and adaptability—skills essential for success in today's fast-paced world. Below is a deeper exploration of some of the most impactful and innovative approaches to learning.

Active learning shifts the focus from passive reception of information to active engagement with the material. Instead of listening to a lecture, students are encouraged to participate through activities that involve analyzing, discussing, or solving problems. This method has shown to improve retention and understanding because students actively work with the concepts they are learning.

- **Examples of Active Learning Strategies:**

- **Think-Pair-Share:** Students think about a question individually, discuss it with a partner, and then share their insights with the class. This approach helps students process their thoughts and builds communication skills.
- **Case Studies:** By examining real-world examples, students apply theoretical concepts to practical scenarios, which enhances critical thinking and decision-making abilities.
- **Problem-Based Learning (PBL):** Students are given a complex problem and must work together to find a solution, often without direct guidance. This builds independence, teamwork, and problem-solving skills.

Project-Based Learning is an approach where students work on a project over an extended period, from several days to an entire term. These projects typically involve complex, real-world problems or challenges, and require students to research, plan, and execute a solution. The process of working on a project teaches students a range of valuable skills, such as time management, collaboration, and critical thinking.

- **How PBL Works:**

- **Choosing the Project:** Ideally, projects are chosen based on students' interests, making them more motivated to engage deeply with the topic.
  - **Research and Exploration:** Students must conduct research, often requiring them to go beyond the textbook and explore various sources, such as interviews, online databases, and empirical experiments.
  - **Presenting Solutions:** At the end of the project, students typically present their findings or products to the class, a broader audience, or community members, reinforcing their communication skills and giving them a sense of accomplishment.
- **Benefits of PBL:** Project-based learning allows students to see the relevance of what they are studying and understand its application in the real world. Additionally, PBL fosters a deeper understanding of the subject matter, as students are not just memorizing facts but working through challenges that require them to apply knowledge in practical ways.

Blended learning combines traditional face-to-face instruction with online or digital learning. This approach leverages the strengths of both in-person and digital education, creating a more flexible and adaptable learning environment.

- **Blended Learning Models:**

- **Flipped Classroom:** In this model, students learn new content at home through video lectures, readings, or online modules, and then apply what they've learned in the classroom through activities and discussions with their teacher and peers.
  - **Station Rotation:** In a station rotation model, students rotate through different learning stations, some of which are online, while others involve teacher-led instruction or collaborative work.
  - **Self-Paced Learning:** This model allows students to progress through online modules at their own pace, with teachers available to provide support as needed.
- **Advantages of Blended Learning:**
    - **Flexibility:** Students can learn at a pace that suits them, revisiting materials or moving ahead as needed.
    - **Personalized Feedback:** Online platforms can provide instant feedback, helping students understand their mistakes and make adjustments in real-time.
    - **Improved Engagement:** By incorporating various types of media, such as videos, quizzes, and interactive simulations, blended learning keeps students engaged and makes learning more dynamic.

Gamification integrates game-like elements into learning activities to increase motivation,



engagement, and retention. By incorporating features like rewards, levels, and challenges, gamified learning makes educational content more appealing and can be particularly effective for younger students.

- **Key Components of Gamified Learning:**
  - **Points and Rewards:** Students earn points or badges for completing tasks, which gives them a sense of accomplishment and encourages continued effort.
  - **Levels and Progression:** As students complete tasks, they can “level up” to more challenging content, creating a structured progression that maintains interest.
  - **Competition and Collaboration:** Some gamified learning experiences encourage friendly competition between students or collaborative efforts to reach shared goals.
- **Benefits of Gamification:**
  - **Increased Motivation:** The gaming elements make learning more enjoyable, encouraging students to take on challenges and persist even when the material is difficult.
  - **Immediate Feedback:** In a gamified environment, students often receive immediate feedback, which reinforces learning and enables quick corrections.
  - **Skill Building:** Many educational games emphasize skills like problem-solving, strategic thinking, and hand-eye coordination, which complement traditional learning objectives.

Experiential learning is based on the principle that students learn best by doing. It emphasizes hands-on experiences, reflection, and real-world applications, allowing students to learn through direct involvement.

- **Examples of Experiential Learning Activities:**
  - **Internships and Apprenticeships:** By working in a real job setting, students gain practical skills and insight into their field of interest.
  - **Service Learning:** This combines academic study with community service, helping students apply their knowledge to help others while gaining an understanding of societal issues.
  - **Simulations and Labs:** Simulations in fields like science, engineering, and healthcare allow students to apply theoretical knowledge in controlled, realistic settings.
- **Advantages:**
  - **Deeper Understanding:** Hands-on experiences help students connect theoretical concepts to real-world applications, deepening their understanding and memory retention.
  - **Development of Soft Skills:** Experiential learning develops skills such as communication, empathy, adaptability, and resilience.
  - **Enhanced Career Readiness:** By gaining real-world experience, students are better prepared for professional roles, with practical skills that can be immediately applied.

Social and collaborative learning is based on the idea that learning is a social process. By engaging in discussions, group work, and peer-to-peer interactions, students can deepen their understanding of concepts, clarify misunderstandings, and build interpersonal skills.

- **Collaborative Learning Techniques:**
  - **Peer Teaching:** Students take turns teaching each other, which reinforces their own understanding and encourages cooperative learning.
  - **Group Projects:** These projects require students to work together towards a common goal, often requiring them to plan, delegate tasks, and resolve conflicts.
  - **Discussion Forums:** Online or in-person discussion forums enable students to share insights, ask questions, and learn from each other’s perspectives.
- **Benefits of Social and Collaborative Learning:**
  - **Enhanced Critical Thinking:** By discussing different viewpoints, students develop critical thinking skills and learn to evaluate arguments.

- **Increased Engagement:** Working with peers can make learning more enjoyable and motivating.
- **Skill Building:** Collaboration teaches students essential social skills, such as communication, empathy, and teamwork.

Inquiry-based learning encourages students to ask questions and seek answers through investigation, exploration, and research. This approach fosters curiosity and encourages a proactive approach to learning, empowering students to become self-motivated, lifelong learners.

- **Stages of Inquiry-Based Learning:**

- **Questioning:** Students start by asking questions about a topic or problem.
- **Research and Investigation:** They gather information, conduct experiments, or engage in discussions to explore potential answers.
- **Solution or Conclusion:** Based on their research, students draw conclusions or solve the problem.
- **Reflection:** Students reflect on their learning experience and the knowledge they gained, reinforcing the inquiry process.

- **Benefits of Inquiry-Based Learning:**

- **Development of Research Skills:** Students learn how to gather and evaluate information, an essential skill for academic and professional success.
- **Increased Curiosity and Independence:** By guiding their own learning, students develop a love for learning and the ability to learn independently.
- **Improved Problem-Solving Abilities:** Inquiry-based learning teaches students to think critically and solve problems logically and creatively.

Adaptive learning is an advanced educational approach that leverages technology to create a customized learning experience for each student. This method uses data-driven algorithms, artificial intelligence (AI), and machine learning to assess individual student performance and adapt instructional content in real time. By doing so, adaptive learning personalizes the pace, path, and presentation of material to match each learner's strengths, weaknesses, preferences, and progress.

### **Core Principles of Adaptive Learning :**

1. **Personalization:** At its core, adaptive learning is about creating a tailored learning experience. The technology behind it collects data on how a student interacts with the content, their accuracy, and speed in answering questions, as well as other factors. This allows the system to present material that suits their current understanding, adjusting complexity, format, or feedback as needed.
2. **Real-Time Adjustment:** Adaptive systems are designed to adjust the learning path based on real-time feedback from students. As they progress through lessons, the software analyzes their responses to continually provide content that aligns with their evolving level of comprehension and ability.
3. **Continuous Assessment:** Rather than relying solely on traditional assessments, adaptive learning platforms conduct continuous assessments. This means the software is constantly monitoring students' interactions and progress, allowing it to identify and respond to learning gaps or misunderstandings immediately.
4. **Data-Driven Insights:** Adaptive learning platforms rely heavily on data analytics to monitor individual and group performance. This data can provide educators with actionable insights into students' progress, highlighting areas where they excel or need additional support.

The adaptive learning process involves several key components that work together to provide a responsive learning experience:

1. **Initial Diagnostic Assessment:** Many adaptive platforms begin with an assessment to gauge the student's starting point. This baseline data helps the system understand the student's initial strengths and areas for improvement, creating a foundation for the personalized learning path.
2. **Content Delivery and Interaction:** Based on the diagnostic results, the system provides tailored content, which could include videos, interactive quizzes, readings, or practice

exercises. As the student interacts with the material, the platform continuously collects data on their performance and engagement.

3. **Real-Time Adjustments:** Adaptive learning systems track metrics such as time spent on each task, accuracy, the number of attempts needed, and patterns of responses. Using this information, the platform can adapt the level of difficulty, suggest additional resources, or offer hints when a student is struggling.
4. **Ongoing Feedback and Support:** Students receive immediate feedback, which is essential for effective learning. If a student answers a question incorrectly, the system might provide an explanation or guide them through a simpler problem before progressing. This feedback loop ensures that students correct misunderstandings early, building a stronger foundation for future learning.
5. **Progress Monitoring and Reporting:** Adaptive platforms provide both students and educators with insights into performance. Teachers can use these reports to adjust instructional strategies or group students for targeted support.

Adaptive learning offers numerous benefits for students, educators, and institutions alike:

1. **Enhanced Learning Outcomes:** Adaptive learning systems identify gaps in understanding early and adjust content to reinforce foundational concepts. This helps students develop a strong grasp of essential skills before advancing to more complex topics, improving overall learning outcomes.
2. **Individualized Pacing:** One of the most significant advantages of adaptive learning is that students can learn at their own pace. High-performing students can move through material more quickly, while those who need extra time or support aren't left behind. This pacing flexibility accommodates different learning speeds, reducing stress and boosting confidence.
3. **Efficient Use of Time:** Adaptive systems streamline the learning process by focusing on areas where students need improvement. Rather than spending time on content they already understand, students can concentrate on areas that require more attention, making their study time more productive.
4. **Increased Student Engagement:** Adaptive learning platforms often use engaging, interactive content that holds students' attention and makes learning more enjoyable. By adjusting to each learner's level, the system keeps content challenging but achievable, promoting intrinsic motivation.
5. **Immediate Feedback and Correction:** The instant feedback that adaptive systems provide helps students learn from their mistakes right away, improving retention and understanding. Immediate feedback has been shown to enhance learning efficiency, as it prevents the reinforcement of incorrect information and allows students to correct errors promptly.
6. **Teacher Empowerment:** Adaptive learning doesn't replace educators; rather, it enhances their ability to support each student effectively. Teachers receive data insights on individual and group performance, allowing them to address learning gaps, identify trends, and provide targeted instruction or enrichment activities.

Numerous technologies and platforms are available to support adaptive learning. These tools often vary by subject, age group, and educational level but share common features, such as data tracking, interactive content, and AI-powered adjustment algorithms. Some popular adaptive learning platforms include:

1. **ALEKS (Assessment and Learning in Knowledge Spaces):** This platform is widely used in mathematics and science education. ALEKS assesses students' current knowledge, identifies gaps, and provides personalized learning paths to address those gaps.
2. **Knewton:** Knewton's adaptive technology is integrated into various educational products, offering personalized content recommendations and insights based on student performance. Knewton's platform supports a wide range of subjects and is used by educational institutions worldwide.

3. **Smart Sparrow:** This platform focuses on adaptive e-learning in higher education. It allows instructors to create interactive and adaptive courseware, providing students with a customized experience based on their unique needs and progress.
4. **DreamBox Learning:** Designed primarily for K-8 mathematics education, DreamBox combines adaptive learning with gamification to create an engaging learning experience. The system adjusts lessons in real-time to keep students challenged and motivated.

While adaptive learning has great potential, successful implementation requires thoughtful planning and consideration:

1. **Teacher Training:** Teachers need to be familiar with adaptive learning tools to make the best use of the data they provide. Proper training helps educators understand how to interpret analytics, use insights for instruction, and monitor student progress.
2. **Blending Adaptive Learning with Traditional Methods:** Adaptive learning works best when combined with traditional instructional methods. While adaptive platforms can provide individualized practice, teachers still play an essential role in leading discussions, clarifying complex concepts, and offering emotional and motivational support.
3. **Ensuring Access to Technology:** For adaptive learning to be effective, students must have reliable access to technology, whether through school-provided devices or at-home access. Ensuring digital equity is essential to prevent further disparities in learning opportunities.
4. **Continuous Evaluation and Improvement:** Institutions should regularly evaluate the effectiveness of adaptive learning systems and make adjustments as needed. This may involve collecting feedback from students and teachers, reviewing data on learning outcomes, and updating content to align with curriculum standards.

Despite its benefits, adaptive learning also presents several challenges:

1. **Privacy and Data Security:** Adaptive learning platforms collect significant amounts of data on student performance. Schools and institutions must ensure compliance with privacy regulations and implement strong security measures to protect sensitive student information.
2. **Risk of Over-Reliance on Technology:** While adaptive systems are powerful tools, there is a risk of over-relying on technology. Educators must strike a balance between digital instruction and traditional teaching methods to provide a holistic learning experience.
3. **Alignment with Learning Standards:** Adaptive content must align with national and regional educational standards. Institutions must ensure that adaptive platforms cover essential learning objectives and offer a curriculum-aligned experience.
4. **Addressing Different Learning Preferences:** While adaptive systems adjust content based on performance, they may not always account for different learning styles or preferences. Integrating features that support visual, auditory, and kinesthetic learners can help make these systems more inclusive.

As technology continues to advance, the potential of adaptive learning will expand further:

1. **AI-Powered Improvements:** Advances in AI and machine learning will enable even more accurate and personalized learning experiences. Future systems may incorporate more sophisticated algorithms that predict learning paths and provide tailored interventions with increasing accuracy.
2. **Integration with Virtual Reality (VR) and Augmented Reality (AR):** Integrating VR and AR into adaptive learning environments can provide immersive experiences that enhance engagement, particularly in subjects that benefit from visualization, such as biology, history, and engineering.
3. **Cross-Disciplinary Applications:** Adaptive learning may expand to cross-disciplinary applications, helping students connect concepts across different subjects. For example, an adaptive system might help a student struggling with reading comprehension in science by connecting it to their strengths in a different subject, such as language arts.
4. **Greater Accessibility and Inclusivity:** Adaptive learning has the potential to become more inclusive, with features designed to support students with disabilities or learning differences.

Enhanced customization options could help address specific needs, providing an equitable learning experience for all.

#### 4. CONCLUSION

Integrating adaptive and personalized learning represents a significant advancement in educational practice, enabling a tailored approach that can adapt to the unique needs and progress of each student. Together, these methods foster a supportive, efficient, and engaging learning environment that nurtures student potential and prepares them for future academic and career success. As technology continues to evolve, the possibilities for combining adaptive and personalized learning will only expand, paving the way for an increasingly inclusive and effective educational landscape. Innovative approaches to learning are reshaping education by placing students at the center of the learning process. Each method, from active and experiential learning to gamification and inquiry-based strategies, offers unique benefits that can be adapted to different learning environments and goals. By combining these approaches, educators can create a rich, diverse learning experience that prepares students not just for exams, but for a lifetime of learning and personal growth. In the future, these methods are likely to become even more integral to education, helping students develop the skills they need to thrive in an increasingly complex world. Adaptive learning is revolutionizing education by personalizing the learning journey for each student. This approach empowers learners to progress at their own pace, providing just-in-time support and feedback that fosters understanding and confidence.

#### References

1. Brusilovsky, P., & Millán, E. (2007). "User Models for Adaptive Hypermedia and Adaptive Educational Systems." In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization*. Springer-Verlag.
2. McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., & Mumper, R. J. (2014). "The Flipped Classroom: A Course Redesign to Foster Learning and Engagement in a Health Professions School." *Academic Medicine*, 89(2), 236–243.
3. Siemens, G., & Long, P. (2011). "Penetrating the Fog: Analytics in Learning and Education." *EDUCAUSE Review*, 46(5), 30–32.
4. Fischer, C., Fishman, B., Dede, C., Eisenkraft, A., & Frumin, K. (2018). "Investigating How Educators' Values and Beliefs Influence Personalized Learning Practices in the Context of a Research–Practice Partnership." *Educational Technology Research and Development*, 66(5), 1231–1254.
5. OECD (2019). "Trends Shaping Education 2019." OECD Publishing.
6. Roschelle, J., Murphy, R. F., Feng, M., & Mason, C. A. (2016). "Digital Formative Assessment for K–12 Personalized Learning." SRI Education.
7. Bishop, J. L., & Verleger, M. A. (2013). "The Flipped Classroom: A Survey of the Research." *Proceedings of the ASEE Annual Conference & Exposition*.
8. Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). "Continued Progress: Promising Evidence on Personalized Learning." RAND Corporation.

**Pravdić Predrag** (*Academy od Professional Studies, Department in Kruševac, Serbia*), **Violeta Đorđević**, **Jelena Erić-Obućina**, **Snežana Gavrilović** (*Academy od Professional Studies, Department in Trstenik, Serbia*)

## PERSONALIZED EDUCATIONAL PROGRAMS WITH AI

***Abstract:** Personalized educational programs are designed to meet the unique learning needs, goals, preferences, and pace of each student. Unlike traditional "one-size-fits-all" methods, these programs use individualized instruction strategies, adaptive technologies, and often data analytics to cater to the distinct requirements of every learner. As technology and educational research continue to evolve, the future of personalized learning looks promising. Advances in AI, machine learning, and data analytics will enable even greater customization, creating highly responsive programs that can address each student's unique learning profile.*

***Key words:** personalized educational programs, teaching, studies*

### 1. INTRODUCTION

Personalized educational programs are designed to meet the unique learning needs, goals, preferences, and pace of each student. Unlike traditional "one-size-fits-all" methods, these programs use individualized instruction strategies, adaptive technologies, and often data analytics to cater to the distinct requirements of every learner. The ultimate goal is to help students achieve their potential by making education more relevant, engaging, and effective. Personalized educational programs have the potential to transform learning for individuals across the lifespan, offering a pathway to a more inclusive, engaging, and effective educational system. As the field of education continues to embrace personalization, it promises a future where every learner's potential can be unlocked and nurtured. The rise of personalized educational programs signals a profound shift in how we understand and approach learning. Traditional education systems, which largely focus on uniform curricula and standardized assessments, often struggle to meet the diverse needs of individual learners. Personalized education aims to address this by creating tailored pathways that consider each student's unique abilities, interests, pace, and goals. This shift has the potential to close educational gaps, particularly for underserved or struggling students, by offering a level of attention and adaptation previously unimaginable in conventional settings. Key to the success of personalized education are the technological advancements that drive it. Through adaptive learning systems, students receive real-time feedback and customized content that helps them engage deeply with material, learn more efficiently, and achieve mastery at their own pace. The incorporation of data analytics and AI allows for continuous adjustments, predicting learning challenges before they escalate and enabling targeted interventions. Virtual and augmented reality add immersive experiences, bringing subjects to life and making complex concepts easier to grasp, while gamification and interactive simulations transform the learning process into an engaging journey. One of the most profound impacts of personalized education is its focus on fostering lifelong learners. By nurturing critical skills such as self-directed learning, problem-solving, and resilience, these programs prepare students not just academically, but for real-world success. Learners who experience personalized education often gain greater self-awareness and independence, learning to set and achieve personal goals. This adaptability is essential in a world where industries are constantly evolving, and where the ability to learn and relearn is invaluable.

### 2. COMPONENTS OF PERSONALIZED EDUCATIONAL PROGRAMS BY AI

Key Principles of Personalized Educational Programs :

1. **Learner-Centered Approach:** Personalized programs place the student at the center of the learning experience. Each learner's interests, abilities, and career aspirations shape their educational path, making learning more relevant and motivating.

2. **Flexible Curriculum:** These programs are often flexible, allowing students to explore subjects of interest or engage in different levels of difficulty based on their progress and preferences. The curriculum may be modular or competency-based, giving students the freedom to work on topics at a pace that suits them.
3. **Adaptive and Differentiated Instruction:** Teachers and educational platforms provide differentiated instruction based on individual student needs. Adaptive technology plays a significant role here, offering content adjustments and real-time responses to students' strengths and weaknesses.
4. **Goal-Oriented Learning:** Personalized educational programs are structured around specific, measurable learning goals. Students are often actively involved in setting these goals, which helps foster accountability, self-direction, and motivation.
5. **Continuous Feedback and Assessment:** Ongoing assessments are used to track students' progress and adjust learning paths accordingly. Frequent feedback allows students to understand their progress, identify areas for improvement, and stay motivated throughout their educational journey.
6. **Student Autonomy:** Personalized programs encourage students to take ownership of their learning. Students may have the option to choose how they learn (e.g., through hands-on projects, digital resources, or collaborative assignments), which fosters independence and critical thinking.

Creating an effective personalized program requires a multi-faceted approach. The following components are essential to building a program that truly adapts to each learner's needs:

1. **Individual Learning Plans (ILPs):** Each student has a unique learning plan that outlines their goals, strengths, areas for improvement, and preferred learning methods. ILPs are typically developed with input from both students and educators, ensuring alignment with curriculum standards while allowing for personal choice.
2. **Competency-Based Learning:** In competency-based models, students progress based on mastery of a subject rather than time spent in the classroom. This allows students to move through content at their own pace and ensures a deeper understanding before advancing to more complex topics.
3. **Customized Content and Resources:** Personalized programs provide various types of learning materials tailored to different learning styles and preferences. For example, students who excel in visual learning might receive video-based content, while those who learn better through reading may be offered more text-heavy materials.
4. **Blended Learning:** Blended learning combines traditional face-to-face instruction with online resources, enabling personalized pacing and access to a broader range of materials. Students can review content independently online, while in-person sessions focus on deepening understanding and addressing specific questions.
5. **Project-Based Learning (PBL):** PBL is often integrated into personalized programs to give students the opportunity to explore real-world problems and apply their knowledge in meaningful ways. These projects are usually self-paced and provide flexibility in how students approach their learning.
6. **Data-Driven Insights:** Personalized programs rely on data to monitor student progress and engagement. Data collected from quizzes, assignments, and interaction patterns helps educators and adaptive learning platforms adjust the program, ensuring that each student remains on a path suited to their learning level.
7. **Teacher Support and Mentorship:** Teachers play a vital role in guiding and mentoring students. In personalized programs, teachers often serve more as facilitators, offering targeted support, encouragement, and resources rather than dictating every step of the learning process.

## Benefits of Personalized Educational Programs

1. **Improved Engagement:** Students are more likely to stay engaged when learning aligns with their interests, strengths, and preferred learning styles. Personalized educational programs make learning more relevant, helping students find purpose and joy in their studies.
2. **Greater Academic Achievement:** By focusing on individual needs and learning styles, personalized programs enable deeper understanding and retention of knowledge. Studies have shown that students in personalized settings often achieve better academic outcomes compared to those in traditional models.
3. **Increased Motivation and Confidence:** Setting personal goals and achieving them can boost students' confidence and motivation. When students see that their unique needs are being met, they feel more supported and are more likely to take initiative in their learning journey.
4. **Enhanced Skills for Future Success:** Personalized programs encourage self-directed learning, problem-solving, and time management—skills that are valuable in higher education and the workforce. Students also develop resilience and adaptability as they work through challenges at their own pace.
5. **Reduced Learning Gaps:** Personalized learning enables educators to address learning gaps as soon as they appear. Rather than advancing all students at the same rate, educators can ensure that each student has a solid foundation before moving on to more complex material.
6. **Opportunities for Exploration:** Personalized programs often provide students with the flexibility to explore topics of personal interest. This freedom promotes curiosity and allows students to develop specialized knowledge in areas they are passionate about.

Personalized educational programs have been implemented across various educational institutions and levels. These programs can vary widely depending on available resources, technological infrastructure, and pedagogical goals:

1. **Primary and Secondary Schools:** In settings, personalized learning is often implemented through adaptive software, blended learning models, and differentiated instruction strategies. Teachers may use platforms that track students' progress and provide recommendations for personalized content.
2. **Higher Education:** Many universities are incorporating personalized programs to accommodate diverse student populations. Competency-based courses, self-paced online modules, and individualized mentorship allow students to tailor their learning experiences to their academic and career goals.
3. **Special Education:** Personalized educational programs have significant benefits for students with special needs. Individualized support and customized resources help ensure these students receive the attention they need to succeed in the mainstream classroom.
4. **Corporate Training and Professional Development:** In the corporate world, personalized learning programs help employees develop specific skills needed for their roles. These programs often incorporate microlearning modules, interactive simulations, and on-the-job training to provide relevant and customized learning experiences.

While the benefits of personalized education are substantial, there are also challenges and considerations that institutions must address for successful implementation:

1. **Resource and Technology Requirements:** Developing a personalized program requires access to adaptive technology, digital content, and data analytics tools. Schools and institutions with limited funding may struggle to implement such programs effectively.
2. **Teacher Training and Support:** Educators must be adequately trained to use personalized learning tools and interpret data insights. Teachers may also need to adapt to new roles as facilitators, providing guidance and support rather than delivering traditional lectures.
3. **Equity and Access:** Personalized education requires equitable access to technology and resources. Institutions must ensure that all students, regardless of socio-economic background, have the tools they need to engage with personalized learning.
4. **Privacy and Data Security:** Data collection is essential for personalized learning, but it also raises privacy concerns. Schools and educational platforms must adhere to strict data privacy standards to protect students' information.



5. **Balancing Personalization with Standardization:** While personalization is important, it must be balanced with standardized curricula and assessments to meet educational benchmarks and ensure fairness across the student population.
6. **Motivation and Self-Discipline:** Personalized programs require students to take more responsibility for their learning. However, not all students may have the intrinsic motivation or self-discipline to thrive in a personalized environment, especially if they are accustomed to traditional structures.

As technology and educational research continue to evolve, the future of personalized learning looks promising. Advances in AI, machine learning, and data analytics will enable even greater customization, creating highly responsive programs that can address each student's unique learning profile. Key trends for the future include:

1. **AI-Driven Personalization:** AI-powered systems can analyze vast amounts of data on student performance and behavior to offer even more nuanced personalization. These systems can anticipate learning gaps before they become problems, provide tailored interventions, and help students achieve mastery more effectively.
2. **Integration with Augmented and Virtual Reality:** AR and VR offer immersive, interactive learning experiences that can be tailored to individual needs. Personalized programs may include virtual field trips, interactive science experiments, or historical reenactments, making education more engaging and accessible.
3. **Holistic Student Profiles:** Future personalized programs may create comprehensive profiles that include not only academic data but also social-emotional insights, learning preferences, and career aspirations. This holistic approach can support the development of well-rounded individuals equipped for future success.
4. **Cross-Disciplinary Personalization:** Personalized educational programs will likely become more interdisciplinary, allowing students to apply concepts across subjects and build connections between fields. For example, a student interested in environmental science could explore personalized modules in biology, chemistry, and public policy.
5. **Increased Collaboration and Peer Learning:** While personalized programs focus on individual learning, they can also incorporate collaborative elements. Students may work on group projects with peers who have complementary strengths, fostering teamwork and communication skills.

### **3. METHODOLOGIES THAT DRIVE PERSONALIZED EDUCATION FORWARD BY AI**

Mechanisms of Personalized Education :

1. **Advanced Data Analytics:** One of the cornerstones of personalized education is the use of data analytics to create real-time insights. Every action a student takes on an educational platform—from answering questions to watching videos—is recorded and analyzed. Machine learning algorithms then identify patterns and tailor recommendations to each student. This allows for immediate adjustments in difficulty, pacing, and content type, creating a truly customized experience.
2. **Student Profiling and Learning Styles:** Personalized educational programs often consider each student's learning style. While some students learn best through visual means, others may thrive with hands-on projects or auditory lessons. Advanced personalization systems create unique learning profiles, taking into account learning styles, cognitive abilities, motivation levels, and interests. This profile is then used to offer learning materials that align with each student's preferred style, boosting both comprehension and retention.
3. **Behavioral and Predictive Modeling:** Modern platforms track student engagement levels, attention spans, and even frustration points, making predictive modeling a powerful tool. If a student consistently struggles with a certain type of problem or loses engagement after specific intervals, the program can proactively introduce breaks, change teaching tactics, or provide motivational prompts to keep them engaged. This predictive capacity helps prevent disengagement and burnout, supporting students to stay on track.

4. **Goal-Setting and Reflection Tools:** Effective personalized programs encourage students to set personal goals and reflect on their progress. These platforms often include goal-setting features that allow students to break down long-term objectives into manageable steps, fostering a growth mindset and resilience. Reflection tools, such as digital journals or progress dashboards, let students see how far they've come and adjust goals as they achieve new milestones.

#### Methodologies that Drive Personalized Education Forward :

1. **Flipped Classroom Model:** This approach reverses the traditional teaching structure, where students learn new material at home and apply it in the classroom. For personalized education, the flipped classroom can offer great benefits. At home, students engage with resources tailored to their individual needs, such as video lessons, readings, or quizzes. In the classroom, they then work on assignments, projects, and discussions that help deepen their understanding with guidance from the teacher.
2. **Project-Based Learning (PBL):** In personalized educational programs, PBL enables students to work on real-world problems that align with their interests. This methodology gives students more control over their learning and allows for interdisciplinary exploration. For instance, a project focused on renewable energy could cover science, math, and social studies, making the learning experience more relevant and memorable.
3. **Gamification and Interactive Learning:** Gamified elements are increasingly popular in personalized education. These include achievement badges, points, levels, and rewards that make learning feel like a game. Interactive simulations and scenarios also allow students to apply their knowledge in dynamic ways, reinforcing learning through practice.
4. **Self-Paced Learning:** Personalized programs often use self-paced modules that allow students to progress through the curriculum as they achieve mastery. This differs from traditional programs that move at a uniform pace. Self-paced learning is especially beneficial for students with busy schedules or unique learning needs, enabling them to catch up or accelerate as needed.

#### Advancements in Technology Supporting Personalized Education :

1. **Artificial Intelligence and Machine Learning:** AI and machine learning are at the heart of modern personalized learning systems. AI helps identify individual student needs, adapt content in real time, and offer predictive insights that empower teachers to make data-driven decisions. Machine learning algorithms, for example, can detect when a student might struggle with a topic and provide supplemental resources or recommend intervention strategies.
2. **Natural Language Processing (NLP):** NLP enables more intuitive interactions between students and digital platforms. For instance, a student struggling with a math problem could type a question into a virtual assistant powered by NLP, which can then offer explanations, resources, or even step-by-step solutions. This technology also aids in language learning, where real-time feedback on grammar, pronunciation, and vocabulary is essential.
3. **Virtual Reality (VR) and Augmented Reality (AR):** VR and AR make immersive, interactive learning experiences possible. In personalized programs, students can take virtual field trips, conduct simulated science experiments, or explore historical sites. This type of learning caters especially well to visual and experiential learners, enhancing engagement and comprehension.
4. **Blockchain for Credentialing and Transparency:** Blockchain technology offers secure credentialing and record-keeping, which can help students and institutions maintain accurate, transparent learning records. Blockchain could enable students to carry a portable learning record from one institution or job to another, documenting their specific skills and competencies without relying on traditional transcripts.

#### Real-World Applications and Examples of Personalized Education :

1. **K-12 Education:** Many K-12 schools are piloting personalized programs through partnerships with edtech providers. Programs like i-Ready, DreamBox, and Khan Academy offer customized math and reading pathways for elementary and secondary students. Teachers can

then use the data from these platforms to create targeted interventions for students who are struggling or provide enrichment for those ready for advanced material.

2. **Higher Education:** Universities are implementing competency-based programs where students can progress at their own pace, a model that works well for adult learners or students with varied backgrounds. For example, Western Governors University (WGU) uses a competency-based model that allows students to advance upon mastering the material, not just completing course hours. This method aligns well with personalized education principles and helps working professionals earn degrees on flexible schedules.
3. **Corporate Training:** Personalized learning is valuable in corporate training, where companies like LinkedIn Learning and Coursera for Business offer individualized learning paths to employees. Employees can engage in self-paced courses tailored to their roles, industries, and career aspirations. Companies benefit from employees who are better equipped for specific tasks and professional growth.
4. **Special Education Programs:** Personalized learning has had a significant impact on special education by offering tailored resources and individualized support. Tools like Learning Ally, which provides audiobooks for students with dyslexia, or platforms that use text-to-speech, have made education more accessible for students with disabilities. Personalization allows these students to learn at their own pace and receive materials in formats that suit their learning needs.

#### Benefits and Potential Outcomes :

1. **Enhanced Learning Outcomes:** Studies show that personalized learning can lead to higher academic performance, greater retention, and improved skill mastery. When learning is tailored to their individual needs, students are more likely to grasp complex concepts and retain information over time.
2. **Development of Critical Life Skills:** Personalized educational programs encourage self-management, time management, and problem-solving skills. These abilities not only help students succeed academically but also prepare them for success in the workforce, where adaptability and independence are highly valued.
3. **Greater Inclusivity:** Personalized education models are especially beneficial for diverse student populations. They provide access to tailored resources for students from varied backgrounds, ensuring that each learner has the opportunity to succeed.
4. **Support for Continuous Learning:** Personalized programs can support lifelong learning by providing learners with tools to continue their education beyond formal schooling. This is particularly valuable in today's fast-evolving job market, where continuous skill development is essential.

#### Overcoming Challenges in Implementation :

1. **Equipping Educators:** Teachers and instructors need thorough training to maximize the benefits of personalized education tools. This training should cover data interpretation, use of adaptive technology, and ways to balance individual attention with class-wide objectives.
2. **Ethics and Data Privacy:** Personalized education relies heavily on student data, raising ethical concerns around privacy. Educational institutions and companies must comply with data privacy laws like GDPR and FERPA, ensuring transparent data collection and secure storage.
3. **Balancing Customization with Core Standards:** While personalization is key, educators must still ensure that students meet core learning standards and benchmarks. Customized programs should be designed to maintain rigor while accommodating diverse learning needs.
4. **Resource Allocation and Infrastructure:** Implementing personalized programs requires a substantial investment in technology, infrastructure, and support staff. Schools and institutions, particularly in underfunded areas, may struggle to provide equitable access to these resources.

The future of personalized education will likely include:

- **Global Access and Scalability:** Advances in cloud technology and mobile learning platforms could bring personalized education to a global audience, including remote or underserved communities.
- **Greater Integration of AI:** Future developments in AI may enable even more precise personalization, with systems capable of understanding students' emotional states, learning fatigue, and levels of engagement.
- **Expanding Learning Ecosystems:** Personalized learning could expand beyond classrooms into lifelong ecosystems where learning is woven into all aspects of life, from workplaces to community spaces.
- **Collaborative Personalization:** Future personalized programs might incorporate collaborative elements where students work in groups on customized projects, building interpersonal and collaborative skills in addition to individual mastery.

#### 4. CONCLUSION

Personalized educational programs are reshaping the future of education by prioritizing individual learning needs and empowering students to take control of their education. These programs offer numerous benefits, from improved academic outcomes and student motivation to the development of crucial life skills. By addressing each learner's unique profile, personalized programs make learning more meaningful, equitable, and effective. As technology advances and educators refine their approaches, personalized learning will continue to expand, creating transformative opportunities for students worldwide. In conclusion, personalized educational programs are reshaping the future of learning by putting the individual at the center of the educational journey. Through innovative technologies like AI, machine learning, and data analytics, these programs are adapting to diverse needs, abilities, and learning preferences, creating more inclusive and effective educational environments. They encourage self-paced learning, critical thinking, and self-management skills, all of which are essential in today's fast-paced and ever-evolving world. While challenges such as ethical concerns, resource disparities, and the need for educator training remain, the benefits of personalized education are clear. When implemented thoughtfully, these programs not only enhance academic outcomes but also equip students with skills crucial for success beyond the classroom. As the field continues to advance, personalized education holds the promise of a future where every learner, regardless of background or ability, has the opportunity to thrive in a truly customized and supportive learning environment. This evolution signifies a shift from a one-size-fits-all model to one that values and fosters individual growth, opening doors for a more equitable and empowered society. Beyond academic settings, personalized learning has found valuable applications in corporate and professional development, allowing individuals to stay competitive in the job market by continuously upgrading their skills. In fields like healthcare, engineering, and finance, where ongoing education is crucial, personalized learning pathways enable professionals to acquire specific knowledge and skills relevant to their roles, which in turn supports organizational growth and innovation. Despite its promise, personalized education faces challenges, including concerns over data privacy, the need for equitable access to technology, and the preparation of educators to effectively leverage personalized tools. Addressing these issues requires a multi-faceted approach, from policy reforms to investments in infrastructure and teacher training. Schools and organizations must ensure that personalization efforts protect student data, maintain rigorous standards, and provide universal access to high-quality resources.

Looking to the future, the evolution of personalized education will likely be shaped by further technological advancements and a growing emphasis on learner-centered approaches. As educators, institutions, and governments recognize the long-term benefits of personalization, there will be greater investment in creating scalable, secure, and inclusive solutions that make this approach accessible to learners everywhere. In time, the principles of personalized education—flexibility, inclusivity, and student empowerment—may redefine education across all levels, creating a more adaptive, resilient, and knowledgeable society. Ultimately, personalized education is more than a trend; it's a transformative approach that aligns education with the complexities and opportunities of

the modern world. By embracing this shift, we can foster a generation of learners who are not only academically proficient but also equipped with the skills, confidence, and mindset needed to thrive in an ever-changing global landscape.

### References

1. Watson, J., & Watson, S. L. (2018). "Blended Learning: The New Normal and Emerging Technologies." *International Journal on E-Learning*, 17(1), 7-15.
2. Huang, R. H., & Spector, J. M. (2019). "Emerging Technologies in Personalized Learning." In *Handbook of Research on Emerging Perspectives in Educational Technology* (pp. 170-189). IGI Global.
3. Liu, M., & Lee, Y. H. (2018). "Adaptive Learning Systems: Technologies and Applications." *Computers & Education*, 124, 131-143.
4. Shaffer, D. W., & Squire, K. D. (2014). "Epistemic Games: Designing and Learning through the Practices of Professional Disciplines." *Educational Psychologist*, 49(3), 221-238.
5. Baker, R. S., & Yacef, K. (2009). "The State of Educational Data Mining in 2009: A Review and Future Visions." *Journal of Educational Data Mining*, 1(1), 3-17.
6. Mayer, R. E., & Moreno, R. (2003). "Nine Ways to Reduce Cognitive Load in Multimedia Learning." *Educational Psychologist*, 38(1), 43-52.
7. Morrison, K. (2019). "Personalized Learning and the Shift Toward Competency-Based Education." *American Journal of Distance Education*, 33(3), 161-174.
8. Anderson, C. A., & Dill, K. E. (2000). "Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life." *Journal of Personality and Social Psychology*, 78(4), 772-790.

**Sergiy Kovalevskyy** (Donbas State Engineering Academy, Kramatorsk - Ternopil, Ukraine),  
**Predrag Dasic** (Engineering Academy of Serbia (IAS), Belgrade, Serbia)

## NEURAL NETWORK BASIS OF INNOVATION.

**Abstract:** *The article is devoted to the analysis of the role of neural network technologies in the implementation of innovations across various sectors of the economy and society. The potential of neural networks for process automation, big data analysis, and trend forecasting is explored. The practical application of neural networks in industry, education, science, and economics is highlighted, including forecasting technical failures, optimizing production processes, personalizing education, and fostering innovative economies. Special attention is given to the prospects of neural network integration within the frameworks of Industry 4.0 and Industry 5.0 concepts, focusing on human-centered approaches and sustainable development. The article also emphasizes the need to establish integrated research, educational, and industrial centers, intersectoral platforms, and innovative educational programs to create synergy between education, science, and economics. Proposed models of neural network use aim to adapt the economy to limited resource conditions and develop financial mechanisms to support innovation.*

**Keywords:** *neural networks, Industry 4.0, Industry 5.0, automation, innovation, education, science, economy, sustainable development.*

### 1. Introduction

Neural network technologies are among the most promising tools for implementing innovations across various economic sectors in today's world. Their potential is immense, as they can radically transform approaches to solving strategic tasks, providing a new level of efficiency, precision, and adaptability. These technologies are now becoming the foundation for automation, data analysis, and trend forecasting, which is crucial for many economic sectors, especially amid rapid transformation.

One of the key areas of neural network application is predicting technical failures in complex equipment. For example, in metallurgical complexes, heavy machinery plants, or high-precision equipment, failures can cause significant losses. Neural networks enable effective real-time monitoring of equipment parameters. Convolutional neural networks analyze sensor data, such as vibrations, temperature, or pressure, identifying even the slightest deviations. Recurrent neural networks, particularly LSTMs, recognize patterns that may precede failures, allowing for maintenance planning to minimize risks and costs. In industries such as aviation, these systems can detect potential engine problems before they occur, ensuring uninterrupted operations and substantial resource savings.

Another critical aspect is the automation of production lines. In complex multi-stage processes, such as electronics or automotive manufacturing, neural networks allow for real-time optimization of technological parameters, adapting them to changing conditions. They not only maintain a consistent level of product quality but also learn during operation. For instance, reinforcement learning methods enable production systems to self-adapt to equipment wear or changes in raw material properties. Additionally, the use of convolutional neural networks for image analysis facilitates automated quality control by detecting even the smallest defects in products.

The role of neural networks extends beyond production processes. In marketing and supply chain management, these technologies provide tools for analyzing large datasets from sources such as online sales or social media. This enables forecasting consumer behavior, optimizing inventory, and even creating personalized offers. For example, transformers like GPT and BERT help create interactive customer experiences, increasing loyalty and boosting sales. For industrial enterprises, this translates into the ability to quickly adapt production lines to new market demands, which is particularly important in today's dynamic market environment.

The restoration of Ukraine's industrial, educational, and scientific potential can also be based on neural networks. In industry, they enable the modernization of outdated equipment, the implementation of digital twins of technologies, and increased productivity. In education, the

development of adaptive learning platforms will facilitate the preparation of specialists for high-tech industries. In science, the integration of big data into research processes will accelerate fundamental discoveries.

## **2. Conceptual Foundations of Neural Network Innovations**

Thanks to their ability to analyze large volumes of data, uncover hidden patterns, and make optimal decisions swiftly, neural networks are increasingly becoming indispensable tools in modern industries. The manufacturing sector, which frequently faces challenges such as equipment failures or the need to optimize processes, particularly benefits from neural networks. For instance, models based on recurrent neural networks (LSTM) can predict technical failures, minimizing downtime risks. Neural networks also enable real-time analysis of sensor data, detecting anomalies and suggesting corrective actions. This reduces costs and improves equipment reliability, which is critically important for enterprises focused on high productivity.

The use of robotic systems driven by deep learning algorithms accelerates research and ensures accurate results. This is particularly relevant for tasks such as discovering new materials, where machine learning combined with high-precision experimental setups significantly reduces the time and resources required to achieve scientific breakthroughs.

The educational sector is also actively implementing neural network technologies. Adaptive learning systems based on transformers, such as GPT and BERT, create personalized programs for students, taking into account their needs and individual learning styles. This enhances the efficiency of the learning process and improves outcomes. These systems not only help students master complex technologies but also prepare them for the realities of Industry 4.0 and 5.0 [1,2]

Moving to the concept of Industry 5.0, it should be noted that it emphasizes a human-centric approach. In this context, neural networks serve as instruments for harmonizing the interaction between humans and machines. They provide operators with decision-making support, enable personalized solutions based on individual needs, and even adapt to new conditions. For example, in manufacturing, this can manifest in the development of intelligent energy management systems that reduce costs and ensure environmental sustainability.

Of particular interest is the interplay between science, education, manufacturing, and the economy. The synergy of these domains forms the foundation of innovative development. The results of scientific research are integrated into manufacturing processes through the preparation of qualified personnel, while education programs based on AI help future professionals master the latest technologies. The creation of educational and research platforms, such as digital simulations or virtual laboratories, ensures that students and researchers have access to advanced technologies without significant costs.

## **3. Application of Neural Networks in the Short-Term Perspective**

In the short term, neural networks can significantly transform approaches to construction, resource extraction, metallurgy, mechanical engineering, education, and economics. Their application ensures automation, efficiency, and predictability, which are particularly crucial under conditions of limited resources and high productivity demands.

In construction, neural network technologies provide new opportunities for automating planning and monitoring processes. They dynamically generate project schedules, considering numerous factors such as weather conditions, material availability, or human resources. Neural networks analyze data from drones or surveillance cameras, enabling real-time detection of delays or defects. Special attention is drawn to digital twins that model the behavior of structures, such as bridges or buildings, under various loads, ensuring predictive maintenance and emergency scenario simulations. For instance, a digital twin of a residential complex allows optimization of energy consumption and enhances the comfort of its residents.

In resource extraction, neural networks facilitate intelligent analysis of geological data, identifying new deposits of valuable minerals even in hard-to-reach regions. They also assess the economic feasibility of extraction under resource scarcity. Using predictive models, neural networks

optimize the operation of drilling rigs, reducing energy consumption, and predict potential equipment failures to avoid downtime. For example, in gas extraction, neural networks can automatically adjust drilling parameters based on the rock type, minimizing the risk of accidents.

Metallurgy and mechanical engineering also benefit from the implementation of neural networks. In metallurgy, they optimize temperature regimes in smelting furnaces, enhancing metal quality, and reducing heat losses, thereby ensuring efficient energy consumption. In mechanical engineering, neural networks streamline the automation of design and manufacturing processes. AI-driven generative design proposes optimal structures for components with minimal weight and high strength. Furthermore, neural networks enable precise machining of components, reducing the risk of defects. For instance, in turbine production, neural networks assist in creating optimal blade shapes and ensuring their manufacturing accuracy.

In education, the potential of neural networks allows for the development of personalized learning paths tailored to each student's level of knowledge and needs. Intelligent assessment systems automatically evaluate assignments, analyze errors, and offer recommendations. Adaptive AI-based platforms enable rapid training of specialists in the field of neural networks, utilizing virtual simulators and online courses that adjust to the students' level. This approach is particularly effective for hands-on learning through practical projects that work with real data.

The economy, under resource-constrained conditions, also gains significantly from neural network implementation. They optimize resource allocation by identifying the most efficient budget utilization directions and developing optimal delivery routes considering weather conditions or road status. By analyzing expenses, neural networks can identify weak points and optimize processes. For small businesses, this becomes an opportunity to achieve competitiveness even with limited financial resources.

#### **4. Long-Term Development Prospects**

The long-term development prospects of neural network technologies pertain to the integration between humans and machines, acceleration of scientific research, sustainable production, innovative education, and the management of complex systems. These directions are already forming the foundation for Industry 5.0, which emphasizes a human-centered approach and sustainable development [3-5]. Neural network technologies enable adaptive interfaces capable of intuitively recognizing operator commands while considering their emotional state or physical limitations. For instance, robotic systems equipped with deep learning not only perform tasks but also act as partners for workers, collaboratively solving production challenges. Intelligent exoskeletons provide safety and comfort, while cognitive interaction between humans and robots facilitates the analysis of workers' requests and the provision of immediate solutions.

The scientific domain is also one of the primary beneficiaries of neural network system advancements. These technologies accelerate the modeling of complex physical and chemical processes, replacing traditional mathematical models that previously required years of research. For example, NLP models automate the analysis of scientific publications, highlighting key trends and discoveries. Such systems not only generate new hypotheses but also optimize experimental parameters in real-time, driving breakthroughs in science.

Sustainability and ecology also benefit from the implementation of neural networks. They enable resource optimization by forecasting energy consumption and identifying efficient delivery routes, minimizing transportation costs and emissions. Neural network-based modeling of closed production cycles contributes to waste reduction and the reuse of materials. For instance, in the chemical industry, leftover materials are converted into secondary raw materials, significantly reducing the environmental impact of production.

Digital twins integrated with neural networks open new opportunities in managing complex systems. They continuously analyze data from physical systems, predict potential failures, and develop preventive strategies. Such systems can be used to model intersectoral collaboration, for example, in projects involving universities, factories, and businesses. Intelligent twins develop long-



term development strategies by modeling economic, technological, and educational scenarios, ensuring synergy between these fields.

## **5. Integration and Implementation**

The integration and implementation of neural network technologies require the creation of effective mechanisms that unify education, science, and industry, while also adapting the economy to the constraints of limited resources. Establishing integrated research centers that combine education, scientific research, and industrial needs is a top priority. Such centers should facilitate workforce retraining, the development of innovative solutions, and the rapid transfer of technologies. For example, "Neural Network Innovation Centers" established at universities could offer practical training programs for students and address real-world industrial challenges. This approach would not only create an effective learning environment but also integrate education into industrial processes.

Intersectoral platforms play a vital role in fostering collaboration between various sectors of the economy. These platforms enable the exchange of data and expertise among engineers, researchers, and entrepreneurs, develop standards for implementing AI in industries and transportation, and accelerate startup development. Young teams gain access to real production data, allowing them to test their ideas in conditions as close to practical applications as possible. Such hubs can integrate into global innovation networks, promoting regional initiatives.

Government support for startups is crucial for innovation development. Flexible grant programs should include funding for prototyping, access to equipment through technology parks, and mentorship support. Combining public funding with private investments via incentive programs will encourage businesses to collaborate with young development teams. This will form the foundation for creating competitive technological solutions.

Adapting the economy to the conditions of limited resources is a significant challenge that can be addressed through neural networks. These technologies optimize production cycles, enabling the prediction of energy consumption and effective inventory management. For instance, in metallurgy, neural networks help adjust furnace temperature regimes to minimize energy consumption and predict material needs to ensure their efficient use.

The creation of sustainable financial models to support innovation involves using neural networks for financial flow modeling, risk analysis, and cost optimization. This will allow for more effective allocation of resources between research, implementation, and scaling. Flexible approaches to lending, including government and private programs, will ensure funding access for companies developing cutting-edge technologies.

## **6. Conclusions**

The innovative development of Ukraine is based on the harmonious interaction of education, science, and the economy. These three components form the foundation for sustainable growth and modernization. To ensure this development, it is essential to identify key areas that will enable the creation of an effective innovation ecosystem.

Education should focus on preparing future professionals. This involves the introduction of interdisciplinary programs that combine knowledge in neural networks, robotics, data analysis, and environmental sustainability. Educational institutions must adapt their curricula to meet the demands of the modern labor market, ensuring the deep integration of technologies into the learning process.

Science should shift its focus to solving applied problems, such as forecasting infrastructure conditions, developing new materials, or automating industrial processes. This will enable rapid responses to current challenges and support economic transformation through innovative solutions.

The economy, by integrating new technologies, should become more resilient to global challenges. The transition to a digital economy, optimization of production processes, and implementation of automation are key steps that will enhance international competitiveness.

Neural network technologies play a unifying role in this system. In education, they ensure personalized learning by adapting materials to students' levels. In science, they facilitate the automation of large-scale data analysis and the creation of simulations to predict the development

scenarios of complex systems. In the economy, they optimize resource use, forecast market trends, and support the adaptation of business strategies.

### **Further Research**

Future development requires the creation of innovative models for interaction among education, science, and industry. Platforms for joint projects must be introduced to provide effective coordination among students, researchers, and industrial companies. For instance, virtual laboratories can unite participants to solve applied problems in real time. The use of digital twins for simulating production processes will also become a powerful tool for learning and research.

Innovative educational programs should integrate AI technologies into curricula, enabling learners to acquire practical skills in working with neural networks. The development of dual education programs that combine theory and practice in industrial settings will prepare specialists capable of promptly addressing applied tasks.

Regional scientific and industrial clusters will facilitate the implementation of joint projects between universities, research institutes, and enterprises. This will enable the integration of innovations into industry and the development of a competitive economy.

### **Comprehensive Approaches to Technology Development**

Further technological development requires a systematic approach that considers interdisciplinarity, flexibility, and scalability. Incorporating methods from various fields such as mathematics, computer science, biology, and chemistry will enable the creation of new solutions for pressing challenges. Flexibility in research will ensure rapid adaptation to new challenges through experimental models. Scalability will allow technologies to transition from local to global applications.

For example, using neural networks in biotechnology can accelerate the development of new drugs or materials. Automated forecasting systems will ensure the resilience of industrial systems to crises, supporting their effective operation even under challenging conditions.

Thus, neural network technologies integrated into education, science, and the economy ensure Ukraine's sustainable development. They establish a foundation for an innovative ecosystem that combines intellectual resources, modern technologies, and efficient financial mechanisms. This approach will enable Ukraine to confidently advance toward global technological leadership.

### **References**

1. Kusma V.V., Gerônimo B.M., Zola F.C., Aragão F.V., Chiroli D.M.D.G., Kovalski J.L., 'A maturity model of occupational safety and health in industry 4.0: An analysis in Brazilian organizations', *Journal of Safety and Sustainability*, 2024, ISSN 2949-9267. Available at: <https://doi.org/10.1016/j.jsasus.2024.11.002>.
2. Kans M., Campos J., 'Digital capabilities driving industry 4.0 and 5.0 transformation: Insights from an interview study in the maintenance domain', *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 4, 2024, art. no. 100384, ISSN 2199-8531. Available at: <https://doi.org/10.1016/j.joitmc.2024.100384>.
3. Nikiforidis K., Kyrtoglou A., Vafeiadis T., Kotsiopoulos T., Nizamis A., Ioannidis D., Votis K., Tzovaras D., Sarigiannidis P., 'Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0', *ICT Express*, 2024, ISSN 2405-9595. Available at: <https://doi.org/10.1016/j.icte.2024.12.001>.
4. Adel A., Alani N.H.S., Jan T., 'Factories of the future in industry 5.0—Softwarization, Servitization, and Industrialization', *Internet of Things*, vol. 28, 2024, art. no. 101431, ISSN 2542-6605. Available at: <https://doi.org/10.1016/j.iot.2024.101431>.
5. Liu S., Li P., Wang J., Liu P., 'Toward industry 5.0: Challenges and enablers of intelligent manufacturing technology implementation under the perspective of sustainability', *Heliyon*, vol. 10, no. 15, 2024, art. no. e35162, ISSN 2405-8440. Available at: <https://doi.org/10.1016/j.heliyon.2024.e35162>.

**Vitaliy Yakovyna, Danylo Melnyk** (*Lviv Polytechnic National University, Lviv Ukraine*)

## **AN INTERPRETABLE SYSTEM FOR BREAST CANCER DIAGNOSIS USING XAI METHODS**

**Abstract:** *Breast cancer remains a leading cause of death among women worldwide, requiring early and accurate detection methods. This study aims to improve the interpretability of breast cancer diagnostic systems by integrating Explainable AI (XAI) techniques without compromising prediction accuracy. Utilizing the publicly available InBreast dataset, we employed pretrained EfficientNet and GMIC models finetuned on mammographic images. Integration of XAI methods provided visual explanations of model decisions, enhancing interpretability crucial for clinical practice. The finetuned GMIC model, enhanced with improved regions of interest (ROIs) grouping and filtering methods, achieved an AUC-ROC of 0.857 and a Dice Coefficient of 95.3%, demonstrating high classification and localization accuracy.*

**Keywords:** *Breast Cancer Diagnosis, Interpretability, Explainable AI, Computer Vision, Deep Learning, Convolutional Neural Networks.*

Breast cancer is one of the leading causes of death among women in the USA. Globally, in 2022 alone, 2.3 million patients were diagnosed with breast cancer, with more than 670,000 resulting deaths [1].

One of the primary methods for diagnosing breast cancer is screening mammography. Mammography is a type of X-ray imaging technique that allows radiologists to visualize internal breast structures. Radiologists analyze screening mammograms for tissue abnormalities that may indicate breast cancer. However, according to research [2], the average radiologist achieves a specificity of only 85.2% with a sensitivity of 62.1% in recognizing malignant cases. This performance requires further confirmation of mammography diagnosis through additional methods, such as ultrasound or tomosynthesis. In some cases (10-20% of women with suspicious mammograms), patients are referred for invasive procedures such as ultrasound-guided biopsies, but only 20-40% of these biopsies confirm malignancy [3].

Early and accurate diagnosis is key to improving the quality of life and survival rates of patients. This brings the need for automated, precise, and interpretable diagnostic tools. Despite significant advances in medical technology, AI-based diagnostic methods have not been widely adopted in clinical practice. One of the primary reasons for this low adoption rate is the lack of transparency in AI-based models. In the healthcare domain, interpretability is especially crucial – diagnostic decisions must be explainable to both patients and professionals to build trust and ensure appropriate medical intervention.

The development of automated, accurate, and interpretable diagnostic tools could lead to greater adoption and integration into clinical workflows, ultimately improving diagnostic reliability and patient outcomes.

Breast cancer diagnosis using mammographic images has a significant focus from computer vision researchers. Automated diagnostic systems are vital for early detection, which improves the chances of successful treatment [4]. Existing systems often struggle with poor interpretability of results.

Many studies have aimed to improve classification accuracy using various deep learning techniques and architectures. For example, [5] proposed a system that classifies mammographic images into benign and malignant categories using a YOLO-based model. The researchers utilized the CBIS-DDSM dataset and a private dataset of spot magnification mammograms for training and evaluation. Preprocessing steps were applied to enhance the contrast between calcifications and other tissues, improving the model's ability to detect subtle anomalies.

Similarly, [6] explored classifiers based on pre-trained deep convolutional neural networks such as AlexNet, GoogleNet, and ResNet. They investigated various approaches, including using DCNNs as feature extractors for support vector machine classifiers and the impact of applying principal component analysis to the extracted features. Their method involved cropping regions of interest from images and applying histogram equalization (CLAHE) to improve image quality. The study demonstrated that combining deep learning with traditional machine learning techniques could enhance classification performance on datasets like CBIS-DDSM and MIAS.

In another approach, [7] EfficientNet models were used for classification based on the BI-RADS system with eight categories instead of the common binary classification approach. This method aimed to provide a more detailed assessment aligned with clinical practices, using a private dataset for training and evaluation. However, the use of a private dataset limits the reproducibility and comparison with other studies.

Addressing the issue of interpretability, [2, 8] developed an interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. The system employed multiple convolutional neural networks, starting with a model to identify suspicious regions, followed by patch classification using models like ResNet-34 or ResNet-50. They formulated the task as a multi-label classification problem for the presence of benign and malignant lesions, allowing for the generation of separate saliency maps for visualization. These saliency maps highlight regions of interest, helping clinicians understand the model's decision-making process.

Several limitations of the clinical applicability of existing studies were determined:

- **Use of Private Datasets.** Many of studies utilized private datasets, making it challenging to reproduce results or compare the proposed methods. The lack of publicly available data makes the validation of these models nearly impossible stopping their wider adoption.
- **Insufficient Models Interpretability.** While some models achieve high accuracy, they often do not provide clear explanations of their decisions. This lack of transparency makes it difficult for clinicians to trust and effectively use these systems.
- **Insufficient Validation of Interpretability.** Most studies focus primarily on classification accuracy, ignoring the need for explainability or without the quantitative assessment of interpretability.

These unresolved issues highlight the need for diagnostic models that balance high accuracy with transparency.

Many existing AI models for breast cancer diagnosis function as «black boxes», offering accurate results but little or no insight into how they reach their conclusions. There is limited integration of Explainable AI (XAI) techniques into breast cancer detection models, suggesting opportunities to improve interpretability. Balancing model accuracy with explainability remains a significant challenge. Finding an optimal balance where the model remains accurate while providing understandable explanations is crucial.

The aim of the work is to improve the interpretability of automated breast cancer diagnostic systems by integrating Explainable AI (XAI) techniques without compromising the accuracy of predictions.

To achieve this aim, we will analyze existing approaches and datasets, implement preprocessing methods, develop a system to evaluate interpretability, create an interpretable model, optimize hyperparameters, and validate the proposed approach.

As the main dataset for further research, we selected the public InBreast dataset due to its high-resolution digital mammographic images with detailed annotations. The dataset contains 410

images from 115 cases, including various types of lesions such as masses, calcifications, asymmetries, and architectural distortions. Precise contours of anomalies are provided, manually annotated by radiologists in XML format. The availability and quality of InBreast make it suitable for developing and validating our interpretable diagnostic system.

We utilized convolutional neural networks (CNNs) for the classification of mammographic images. CNNs are effective for image analysis tasks because they can automatically learn hierarchical feature representations. Specifically, we employed pre-trained EfficientNet models [11], known for their balance of accuracy and computational efficiency.

We assessed model performance in terms of both classification and localization quality.

Classification Metrics used for each class:

**Precision** measures the proportion of correct positive predictions:

$$Precision_c = \frac{TP_c}{TP_c + FP_c}.$$

**Recall** measures the proportion of actual positives correctly identified:

$$Recall_c = \frac{TP_c}{TP_c + FN_c},$$

where  $TP_c$  is the number of correct predictions of class  $c$ ,  $FP_c$  is instances incorrectly predicted as class  $c$ , and  $FN_c$  is the number of instances of class  $c$  incorrectly predicted as another class.

We then computed the macro-average over all classes to obtain overall Precision and Recall:

$$Precision = \frac{1}{3} \sum_c Precision_c,$$

$$Recall = \frac{1}{3} \sum_c Recall_c,$$

where  $c \in \{Malignant; Benign; Normal\}$ .

**Area Under ROC Curve** (AUC-ROC) evaluates the model's ability to distinguish between classes across all threshold settings. For per-class AUC-ROC, we used the one-vs-rest method. To aggregate AUC-ROC, we computed:

$$AUC_{roc,agg} = \sqrt{AUC_{roc,Benign} \cdot AUC_{roc,Malignant}},$$

where  $AUC_{roc,Benign}$  is the area under the ROC curve for Benign vs. Rest, and  $AUC_{roc,Malignant}$  is ROC AUC of Malignant vs. Rest.

Used Localization Metrics:

**Dice Coefficient** measures the overlap between the predicted regions and ground truth in terms of pixels:

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}.$$

**Precision and Recall for Localization** show how many Regions of Interest (ROIs) are precisely detected (Precision) and the proportion of actual ROIs correctly identified (Recall) ignoring their areas. A predicted ROI is considered correct if it intersects with annotated anomalies.

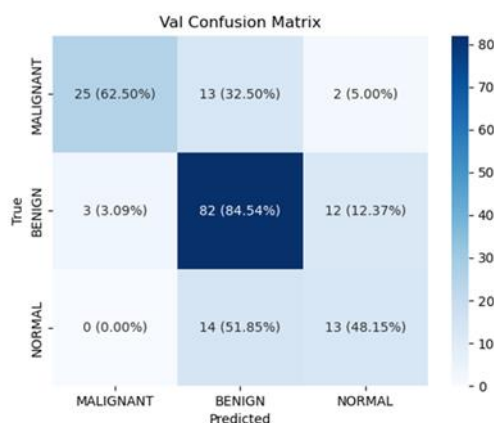
We finetuned EfficientNet-B4 and EfficientNet-B5 models on the InBreast dataset. We experimented with full and partial parameter tuning (training only the final layers while keeping earlier layers frozen). To address class imbalance we applied a weighted sampler, augmenting underrepresented classes in training set.

Table 1. Results of EfficientNet classifiers.

Experiment №	Architecture	Trained Parameters	Balance Train	AUC_ROC_AGG	
				Train	Test
18	EffNet B4	All, 17M	–	0,996	0,834
19	EffNet B4	Freeze, 14M	–	0,999	0,817
21	EffNet B4	All, 17M	+	1,000	0,842
22	EffNet B4	Freeze, 14M	+	0,955	0,805
25	EffNet B5	Freeze, 22M	–	0,965	0,785
26	EffNet B5	Freeze, 22M	+	0,946	0,795

The best classification results were achieved with the EfficientNet-B4 model when all layers were trained and the training set was balanced (Experiment №21): AUC-ROC: 0.842, Accuracy: 73.2%, Precision: 70.9% and Recall: 65.1%.

The model demonstrated better performance in recognizing malignant cases compared to benign ones (Fig. 1). The most common errors were 14 normal images predicted as benign and 13 malignant images predicted as benign.



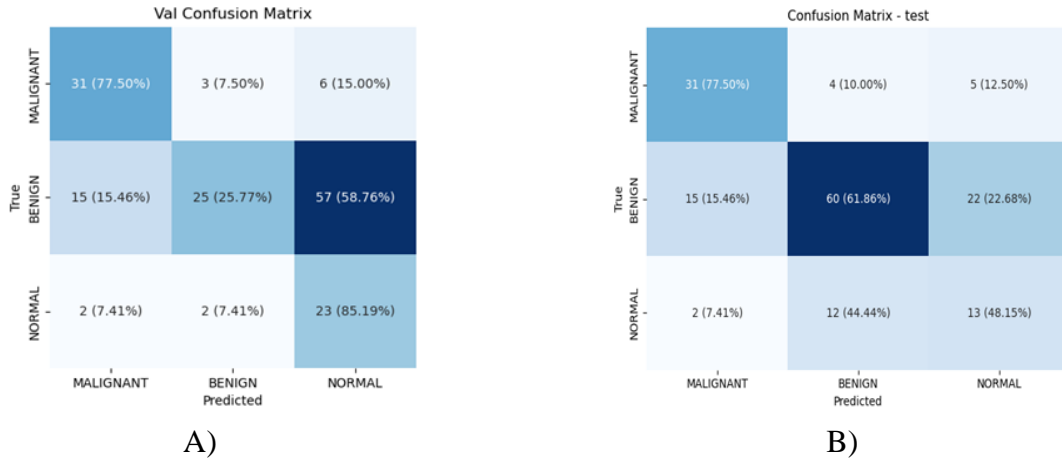
**Figure 1.** Confusion matrix of EfficientNet B4 classifier trained all params with a balanced train set.

We also compared the achieved results with the GMIC model [8] (Table 2).

Table 2. Results of GMIC classifiers.

Experiment №	Architecture	Trained Parameters	Balance Train	AUC_ROC_AGG	
				Train	Test
31	GMIC	No finetuning	–	–	0,502
32	GMIC	Finetune last linear layer, 1K	–	0,872	0,853
33	GMIC	Finetune last linear layer, 1K	+	0,815	0,857

The GMIC model, after finetuning the last linear layer with a balanced train subset (Experiment #33), showed the highest ROC-AUC: 0.857. However, it showed relatively low accuracy after thresholding: ACC: 48.2%. Adjusting threshold values improved classification metrics. Optimizing thresholds to 0.3 for the benign class and 0.5 for the malignant class increased overall accuracy by approximately 17%: ACC: 65.5%, Precision: 58.2%, and Recall: 62.8%. Confusion matrixes before and after thresholds tuning are shown in Fig. 2.



**Figure 2.** Confusion matrix of finetuned GMIC classifier trained with balanced train set. A) Before threshold tuning; B) After threshold tuning.

To enhance the interpretability of the developed classifiers, we integrated Explainable AI (XAI) methods, focusing on:

- **Activation Maps.** These methods visualize the activation levels of different regions by analyzing neuron activations in intermediate network layers, indicating areas of high importance.
- **Class Activation Mapping (CAM).** A technique that identifies the impact of different image regions on the prediction of a specific class. CAM focuses on areas that most contribute to classifying an image into a particular category.
- **Grad-CAM [12] and Variants.** Gradient-weighted Class Activation Mapping (Grad-CAM) and its extensions Grad-CAM++ and HiResCAM were used to generate saliency maps highlighting important regions influencing the model's predictions.

For the EfficientNet classifiers, we used Activation Maps, Grad-CAM, and its variants to generate saliency maps that highlight important regions influencing the model's predictions.

For the GMIC model, we implemented the approach proposed by the authors [8] and developed our own improved method. In the improved method, we applied additional grouping and filtering of predicted regions. Regions close to each other were grouped and approximated with bounding rectangles. Filtering based on confidence scores allowed us to discard insignificant regions.

Explainable AI methods do not always correctly differentiate the severity of anomalies. Therefore, all localization metrics focused solely on the ability to distinguish anomalies from normal regions, ignoring severity levels.

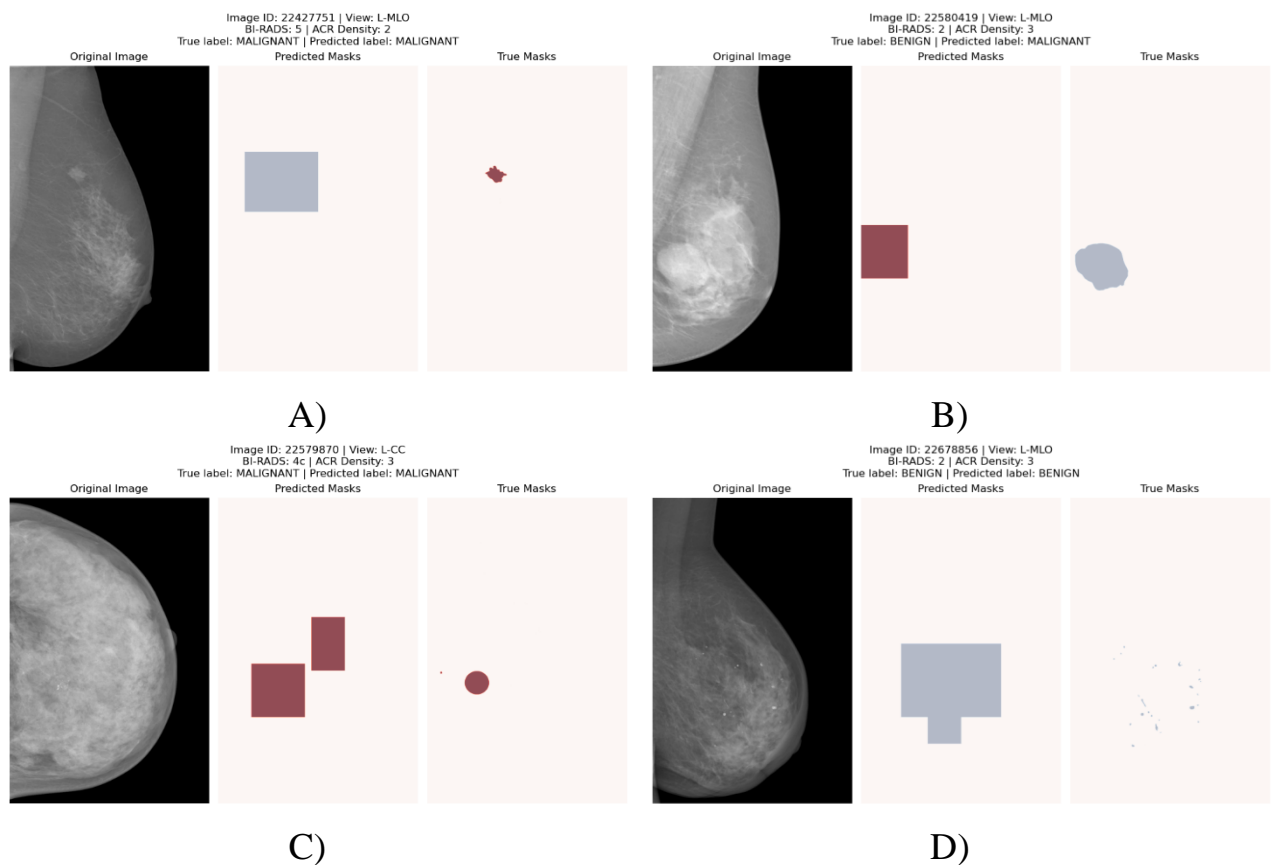
Table 3. Localization metrics on the test set.

Training Experiment №	Classifier	Explainability Approach	Patch Detection Metrics, %		Segmentation Metric, %
			Precision	Recall	Dice
21	EfficientNet B4	Attention maps	75,0	37,8	8,7
21	EfficientNet B4	GradCAM	75,0	37,8	8,7
21	EfficientNet B4	GradCAM++	75,0	37,8	8,7
21	EfficientNet B4	HiResCAM	75,0	37,8	8,7
31	GMIC	GMIC	28,7	35,1	92,9
31	GMIC	GMIC with grouping and	51,5	26,5	94,9

		filtration			
33	Finetuned GMIC	GMIC	41,1	46,2	93,2
33	Finetuned GMIC	GMIC with grouping and filtration.	65,5	37,3	95,3

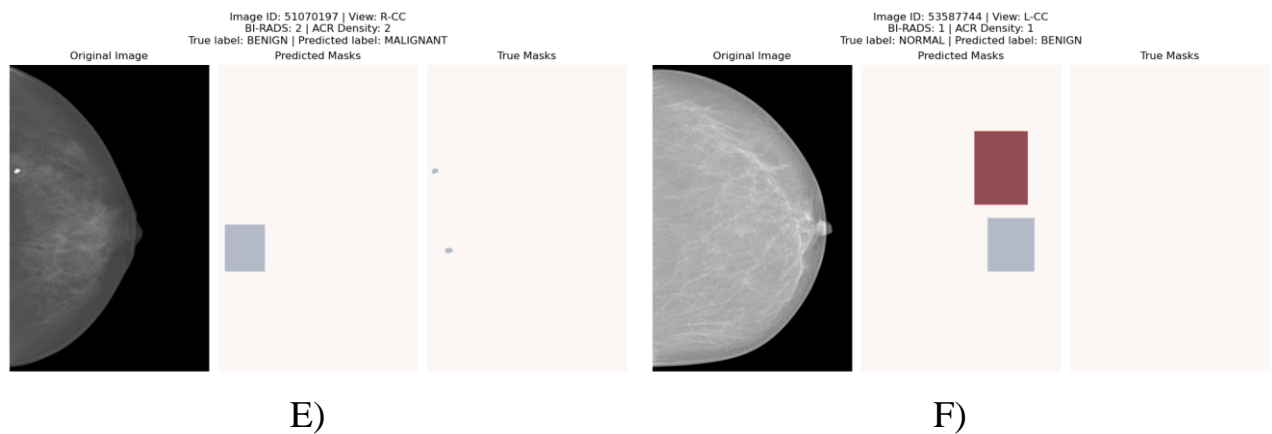
The best localization performance was achieved by the fine-tuned GMIC model with grouping and filtering of anomalous regions (Experiment №33). The patch detection reached a Precision of 65.5% and Recall of 37.3%, with a Dice Coefficient of 95.3%. Let's explore localization results on several examples (Fig. 3):

- (A), (B), (D): Significant differences between the shape of the detected anomaly and the annotated ground truth due to approximating aggregated predictions with simple rectangles.
- (C): A cluster of calcifications was successfully detected and approximated with several rectangles, but some distant calcifications were missed.
- (E), (F): Instances of falsely predicted regions, highlighting the need for better filtering mechanisms.



**Figure 3.** Examples of detection results.





**Figure 3** (continue). Examples of detection results.

## CONCLUSIONS

Our study demonstrates that integrating Explainable AI (XAI) techniques into a convolutional neural network (CNN) models significantly enhances the interpretability of automated breast cancer diagnostic systems using mammographic images. By employing the publicly available InBreast dataset, we developed and fine-tuned models that not only achieve high classification accuracy but also provide visual explanations of their decisions, addressing a critical need in clinical practice.

The fine-tuned GMIC model, combined with our improved method of grouping and filtering predicted regions, achieved the best performance in both classification and localization tasks. Specifically, the model reached an AUC-ROC of 0.857 and a Dice Coefficient of 95.3%. The localization precision and recall were 65.5% and 37.3%, respectively. These results indicate that the model is effective in accurately identifying malignant cases and highlighting the regions of interest that influenced its decisions.

However, challenges remain in fully aligning the detected anomaly shapes with ground truth annotations. Discrepancies often arise due to approximating complex lesion shapes with simple geometric figures like rectangles. Additionally, some distant or subtle anomalies may be missed, and instances of false positives highlight the need for further refinement.

To further enhance the system's clinical applicability, future research should focus on:

**Refining Anomaly Shape Approximation:** Developing more sophisticated methods to approximate the shapes of anomalies can improve the overlap between detected regions and ground truth, leading to better interpretability and trustworthiness.

**Optimizing Threshold Values:** Fine-tuning the threshold settings for region filtering can help balance sensitivity and specificity, reducing false positives and negatives.

**Enhancing Post-processing Techniques:** Implementing advanced image processing methods may improve the quality of localized regions, capturing subtle anomalies that are currently missed and eliminating insignificant regions.

**Validating with Larger and Diverse Datasets:** Testing the models on larger and more diverse datasets can help generalize the findings and ensure the models perform well across different populations and imaging conditions.

By addressing these areas, we aim to develop a more robust and interpretable diagnostic tool that can be confidently integrated into clinical workflows. This advancement has the potential to improve diagnostic reliability, facilitate early detection of breast cancer, and ultimately enhance patient outcomes.

## References

1. WHO Breast Cancer Fact Sheet [Электронный ресурс] / WHO. — 2024. — Режим доступа: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
2. Shen Y. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization / Y. Shen, N. Wu, J. Phang, [et al.] // *Medical Image Analysis*. — 2021. — Vol. 68. — P. 101908.
3. Kopans D. B. An open letter to panels that are deciding guidelines for breast cancer screening / D. B. Kopans // *Breast Cancer Research and Treatment*. — 2015. — Vol. 151, No. 1. — P. 19–25.
4. Duffy S. W. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties: A collaborative evaluation / S. W. Duffy, L. Tabár, H. Chen, [et al.] // *Cancer*. — 2002. — Vol. 95, No. 3. — P. 458–469.
5. Chen J.-L. A YOLO-based AI system for classifying calcifications on spot magnification mammograms / J.-L. Chen, L.-H. Cheng, J. Wang, [et al.] // *BioMedical Engineering OnLine*. — 2023. — Vol. 22, No. 1. — P. 54.
6. Ragab D. A. A framework for breast cancer classification using Multi-DCNNs / D. A. Ragab, O. Attallah, M. Sharkas, [et al.] // *Computers in Biology and Medicine*. — 2021. — Vol. 131. — P. 104245.
7. Tsai K.-J. A High-Performance Deep Neural Network Model for BI-RADS Classification of Screening Mammography / K.-J. Tsai, M.-C. Chou, H.-M. Li, [et al.] // *Sensors*. — 2022. — Vol. 22, No. 3. — P. 1160.
8. Shen Y. Globally-Aware Multiple Instance Classifier for Breast Cancer Screening / Y. Shen, N. Wu, J. Phang, [et al.] // 2019.
9. Singh A. Explainable Deep Learning Models in Medical Image Analysis / A. Singh, S. Sengupta, V. Lakshminarayanan // *Journal of Imaging*. — 2020. — Vol. 6, No. 6. — P. 52.
10. Ennab M. Designing an Interpretability-Based Model to Explain the Artificial Intelligence Algorithms in Healthcare / M. Ennab, H. Mcheick // *Diagnostics*. — 2022. — Vol. 12, No. 7. — P. 1557.
11. Tan M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks / M. Tan, Q. V. Le // 2019.
12. Selvaraju R. R. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization / R. R. Selvaraju, M. Cogswell, A. Das, [et al.] // 2016.

**Афонін Д.І.** (Київський національний університет імені Тараса Шевченка, Київ, Україна)

## МОДЕЛЮВАННЯ ПОШИРЕННЯ ІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ З АГЕНТАМИ

**Анотація.** У даній публікації розглядається проблема моделювання поширення інформації у соціальних мережах. Було розглянуто кілька підходів та методів для цього, але основна увага приділена агентному підходу. Було запропоновано розглянути в подальшому поширення інформації з урахуванням груп у соціальній мережі, запропоновано модель, що здатна проводити таке моделювання та зроблено аналіз отриманих результатів.

**Ключові слова:** поширення інформації, соціальні мережі, моделювання, агентний підхід, групи, запропонована модель, аналіз, методи

**Abstract.** This publication addresses the issue of modeling information diffusion in social networks. Several approaches and methods have been examined, with the primary focus on the agent-based approach. It has been proposed to consider information diffusion in the context of groups within a social network. A model capable of performing such modeling was suggested, and the obtained results were analyzed.

**Keywords:** information dissemination, social networks, modeling, agent-based approach, groups, proposed model, analysis, methods

Соціальні мережі сьогодні відіграють ключову роль у житті суспільства, виступаючи не лише платформами для спілкування, але й потужними засобами поширення інформації та взаємодії між користувачами. Розуміння механізмів поширення інформації в цих мережах є критично важливим для багатьох сфер, включаючи маркетинг, політику, соціальну динаміку та бізнес-стратегії. Однак активний вплив соціальних мереж породжує низку викликів і проблем, таких як поширення фейкових новин, вірусний маркетинг, формування групової думки та поведінкових патернів. Вивчення цих явищ є складним завданням через високу динаміку й складність системи, що включає велику кількість користувачів та їх взаємозв'язків.

Крім того, зростання обсягів даних і розвиток новітніх технологій, таких як штучний інтелект, машинне навчання та аналіз великих даних, створюють нові можливості для вдосконалення математичного моделювання процесів поширення інформації. Це дозволяє підвищити точність і реалістичність моделей, а також глибше зрозуміти зв'язки між факторами, що впливають на поширення інформації.

Відсутність універсальних моделей, які одночасно враховують індивідуальні характеристики користувачів і глобальні властивості мережі, створює науковий і практичний запит на розробку нових підходів до моделювання. Існує потреба у створенні інструментів для передбачення процесів поширення інформації в мережах, боротьби зі спамом, фейковими новинами та негативними впливами та розробки нових стратегій для практичного використання в бізнесі, маркетингу та політиці.

Поширення інформації в соціальних мережах є складним багатофакторним процесом, який потребує детального аналізу та моделювання для розуміння його механізмів і динаміки. Для цього застосовуються різні підходи, такі як статистичні моделі, агентні моделі, соціально-економічні моделі, моделі складних мереж і моделі машинного навчання. Кожен із цих підходів має свої переваги й обмеження.

Зокрема, статистичні моделі дозволяють виявляти загальні закономірності в структурі мережі, але не враховують індивідуальних особливостей користувачів. Агентні моделі надають можливість відтворювати поведінку окремих користувачів і аналізувати їхні стратегії взаємодії, проте вони часто потребують значних обчислювальних ресурсів. Соціально-економічні моделі акцентують увагу на впливі соціальних і економічних факторів, однак можуть бути складними для узагальнення. Моделі складних мереж дозволяють аналізувати глобальну структуру мережі, але їхній фокус на взаємодіях вузлів не завжди дозволяє дослідити поведінку окремих користувачів. Нарешті, моделі машинного навчання

автоматизують аналіз великих наборів даних, однак можуть бути обмеженими у поясненні причинно-наслідкових зв'язків.

Різноманіття підходів відображає складність самого явища. Наприклад, серед агентних моделей окрему увагу привертають такі, як модель конкуренції [1], модель порогового впливу [2], модель широкого поширення [3], модель впливу [4] та модель поширення інновацій [5]. Кожна з них ілюструє специфічні аспекти поширення інформації, проте для моделювання реальних процесів в умовах сучасних соціальних мереж часто недостатньо обмежуватися лише одним підходом.

Відсутність універсальної моделі, яка б враховувала як індивідуальні особливості користувачів, так і глобальні властивості мережі, створює значні виклики для дослідження поширення інформації. Це вимагає розробки комбінованих підходів, які б об'єднували переваги різних моделей, забезпечуючи точність аналізу, врахування соціально-економічних факторів і ефективну обробку великих обсягів даних.

У дослідженні [6] розглянуто та проаналізовано наведені вище моделі та підходи, запропоновано модифікований спосіб передбачення рівнів поширення інформації у соціальній мережі за допомогою агентного підходу. Слід окремо наголосити на важливості виділених груп у соціальній мережі: запропонований метод враховує такий поділ, що дає можливість розглядати процес поширення з урахуванням різних факторів.

Розроблена модель функціонує за принципом ймовірнісного автомата, де всі вузли початково мають атрибут стану, рівний нулю. У разі успішного сприйняття інформації користувач змінює свій стан на одиницю. Для моделювання поширення інформації в соціальній мережі запропоновано наступний алгоритм:

- a) перед початком алгоритму визначаються агенти  $A_{start}$ , які ініціюють поширення інформації;
- b) статус агентів змінюється з нуля на одиницю;
- c) обирається підгрупа користувачів  $V_{prime} \subset V$ , серед яких інформація спочатку поширюється. Також задається критичний поріг  $N_{critical} < N$  після досягнення якого інформація починає розповсюджуватися по всій мережі;
- d) користувачі, які сприйняли інформацію, починають передавати її своїм сусідам зі статусом нуль із певною ймовірністю, водночас:
  - 1) перевіряється, чи належить користувач до групи первинного поширення;
  - 2) ймовірність поширення  $P_{spread}$  зменшується на  $c_{penalty}(g_{number}) \in C_{penalty}$  де  $g_{number}(i) \in S_{group}$ , якщо відправник і отримувач інформації належать до різних груп;
  - 3) виконується спроба передачі інформації;
  - 4) якщо поширення в межах мережі стає неможливим (наприклад, через обмеження кількості ітерацій), алгоритм завершується;
  - 5) процес повертається до кроку 1) та повторюється доти, доки кількість інформованих користувачів не досягне заданого порогу  $N_{critical}$ ;
- e) користувачі, які сприйняли інформацію, починають передавати її всім своїм сусідам зі статусом нуль із певною ймовірністю;
  - 1) ймовірність зменшується, якщо відправник і отримувач належать до різних груп;
  - 2) робиться спроба поширити інформацію;
- f) алгоритм завершується, якщо всі вузли  $V$  змінили свій статус на одиницю;
- g) Якщо є нові вузли, які змінили свій стан, або виконано менше 10 ітерацій, алгоритм повертається до кроку e).

Зважаючи на всі фактори, рекомендується використовувати ймовірність поширення або кількість зв'язків як критерії для визначення користувачів у  $V_{prime}$ . У реальних умовах такими користувачами зазвичай є найпомітніші й найефективніші поширювачі інформації. Крім того, значення  $N_{critical}$  може визначатися залежно від кількості користувачів у  $V_{prime}$ .

Проведено моделювання для двох соціальних мереж з 2000 користувачів та 5000 користувачів з різною кількістю агентів та їх розташування у соціальній мережі. Проаналізовано поширення в кожній окремій групі мережі. Нижче на Рис.1, Рис.2 та Рис.3 наведено деякі результати поширення у мережі на 5000 вершин з одним початковим агентом, зі ста агентами розташованими по всій мережі та зі ста агентами зосередженими в одній групі.

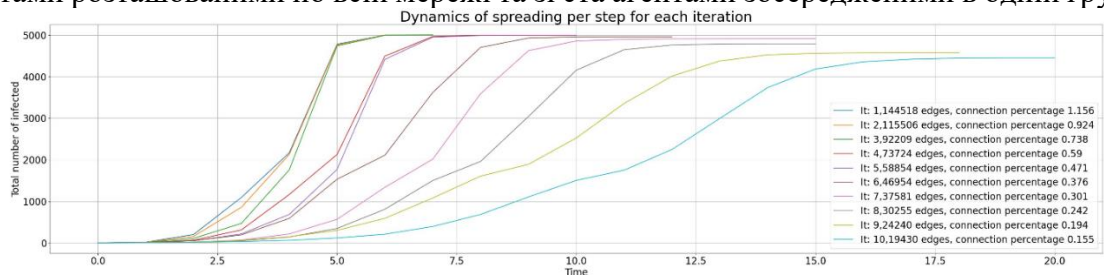


Рис.1 – Динаміка поширення з одним агентом впливу

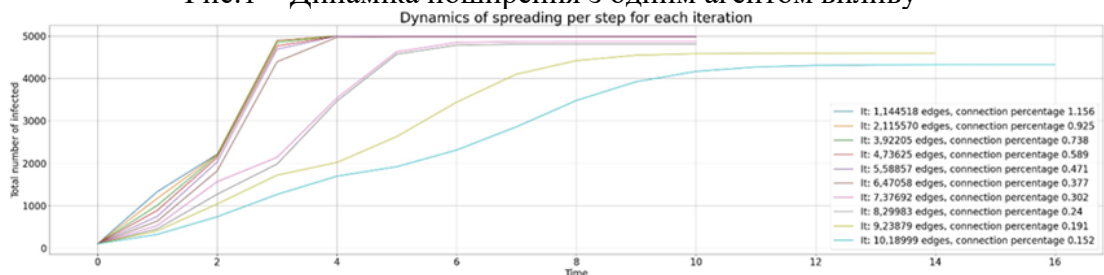


Рис.2 – Динаміка поширення із ста агентами впливу

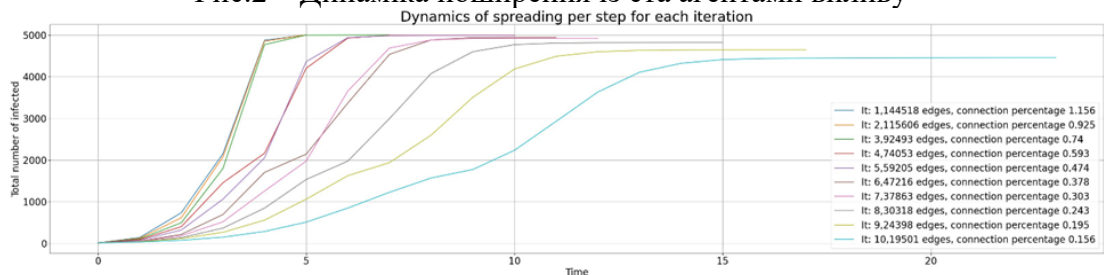


Рис.3 – Динаміка поширення із ста агентами впливу в одній групі

## ВИСНОВКИ

Одним із ключових аспектів моделювання поширення інформації в соціальних мережах є поділ мережі на групи, що відображає її природну структуру. Такий поділ може базуватися на різних критеріях, наприклад, мова спілкування, спільні інтереси (спорт, ігрова індустрія, наука тощо) або інші характеристики. Моделювання поширення інформації в межах кожної групи та між ними дозволяє спростити задачу, перетворюючи її з індивідуального аналізу для кожного агента на процес оцінювання динаміки поширення для цілих груп залежно від типу інформації.

Перспективним напрямком подальших досліджень вбачаю розробку моделей, які враховують особливості взаємодії між групами, а також оптимізація методів і параметрів для точного моделювання поширення інформації. Серед потенційних підходів до розгляду — модель клітинних автоматів, модель агрегації обмеженої дифузії, модель Петрі та нейронні мережі. Аналіз та порівняння цих моделей можуть сприяти виявленню можливостей для їх покращення та адаптації до специфіки досліджуваної проблематики. Це, у свою чергу, відкриває перспективи створення точної, швидкої та ефективної моделі прогнозування поширення інформації в соціальних мережах.

## Список посилань.

1. Huang, Y., Thornton, K. M., & Efthimiadis, E. N. (2015). Competition in social media among heterogeneous agents. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*.

2. *Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*
3. *Centola, D. (2010). The spread of behavior in an online social network experiment. Science*
4. *Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*
5. *Guidolin M. (2023). Innovation Diffusion Models: Theory and Practice*
6. *Afonin D. (2024) Modeling information diffusion processes in social networks with agents*

Віт Р.В., Мазурець О.В. (Хмельницький національний університет, м. Хмельницький, Україна)

## ТЕМАТИЧНА КЛАСИФІКАЦІЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ

**Анотація:** Запропоновано підхід до тематичної класифікації текстової інформації засобами обробки природної мови для автоматизованого ідентифікування та групування текстів за основними темами. Цей підхід сприяє ефективній організації збереження і використання великих обсягів текстової інформації, організуючи структурований доступ до її змісту. Проведена крос-валідація продемонструвала результат точності 0.86, що на 0.15 перевищує точність, отриману при використанні LDA без додаткових модифікацій для класифікації.

**Ключові слова:** тематична класифікація, обробка природної мови, LDA, ідентифікування текстів, групування текстів.

**Abstract:** Approach for thematic classification of text information using natural language processing tools for automated identification and grouping of texts by main topics is proposed. This approach contributes to the effective organization of storage and use of large volumes of text information, organizing structured access to its content. The cross-validation demonstrated an accuracy result of 0.86, which is 0.15 higher than the accuracy obtained when using LDA without additional modifications for classification.

**Keywords:** thematic classification, natural language processing, LDA, texts identification, texts grouping.

### Постановка проблеми

Тематична класифікація текстів є поширеним підходом до обробки та аналізу неструктурованих і напівструктурованих даних в організаціях [1]. Цей процес полягає у групуванні текстової інформації за певними категоріями чи темами, що дозволяє виявляти ключові ідеї, тенденції та шаблони в даних.

Застосування алгоритмів машинного навчання дає змогу автоматизувати аналіз текстів, використовуючи контекстуальні ознаки, що значно підвищує швидкість і точність класифікації.

### Аналіз останніх публікацій

Сучасні огляди часто фокусуються на імовірнісних підходах до тематичного моделювання, але важливо також враховувати методи, які базуються на лінійній алгебрі, оскільки вони здатні ефективно представляти тематичну структуру текстів [2].

У сфері тематичного аналізу текстів, заснованого на машинному навчанні, проведено чимало досліджень, орієнтованих на виявлення ключових слів і фраз, а також на формування n-грам за критерієм релевантності. Наприклад, одне з досліджень використало попередньо навчену модель BERT NLP від SberDevices, адаптовану до російськомовних текстів. Результати свідчать, що цей підхід ефективний для аналізу текстів, якщо тематика добре репрезентована у навчальному наборі даних [3].

Ще одне дослідження вивчало вплив природних криз на функціонування ланцюгів постачання, використовуючи дані соціальних мереж. Для цього була розроблена структура, що дозволяє автоматично оцінювати вплив криз, таких як пандемія COVID-19. Використовуючи аналіз термінів спільного входження та побудову карти знань, дослідники проаналізували 1024 онлайн-звіти. Було визначено п'ять ключових напрямів впливу на ланцюги постачання: роздрібна торгівля продуктами, харчові послуги, виробництво, поведінка споживачів та логістика. Ця модель стала ефективним інструментом для підтримки прийняття рішень у кризових умовах [4].

### Мета роботи та постановка завдань

Метою роботи є розробка методу тематичної класифікації текстової інформації засобами обробки природної мови, здатного підвищити точність і релевантність тематичного аналізу, що сприятиме прийняттю обґрунтованих рішень на основі текстових даних.

### Виклад основного матеріалу

Метод тематичної класифікації текстової інформації засобами обробки природної мови дає змогу перетворювати вхідні текстові дані у вивід у вигляді кількості тем, домінуючої теми

кожного документа та розширеного списку ключових слів для кожної з тем (рис. 1). Розроблений підхід поєднує гнучкість тематичного моделювання з можливістю автоматичного розширення ключових слів, забезпечуючи ефективний тематичний аналіз текстів.

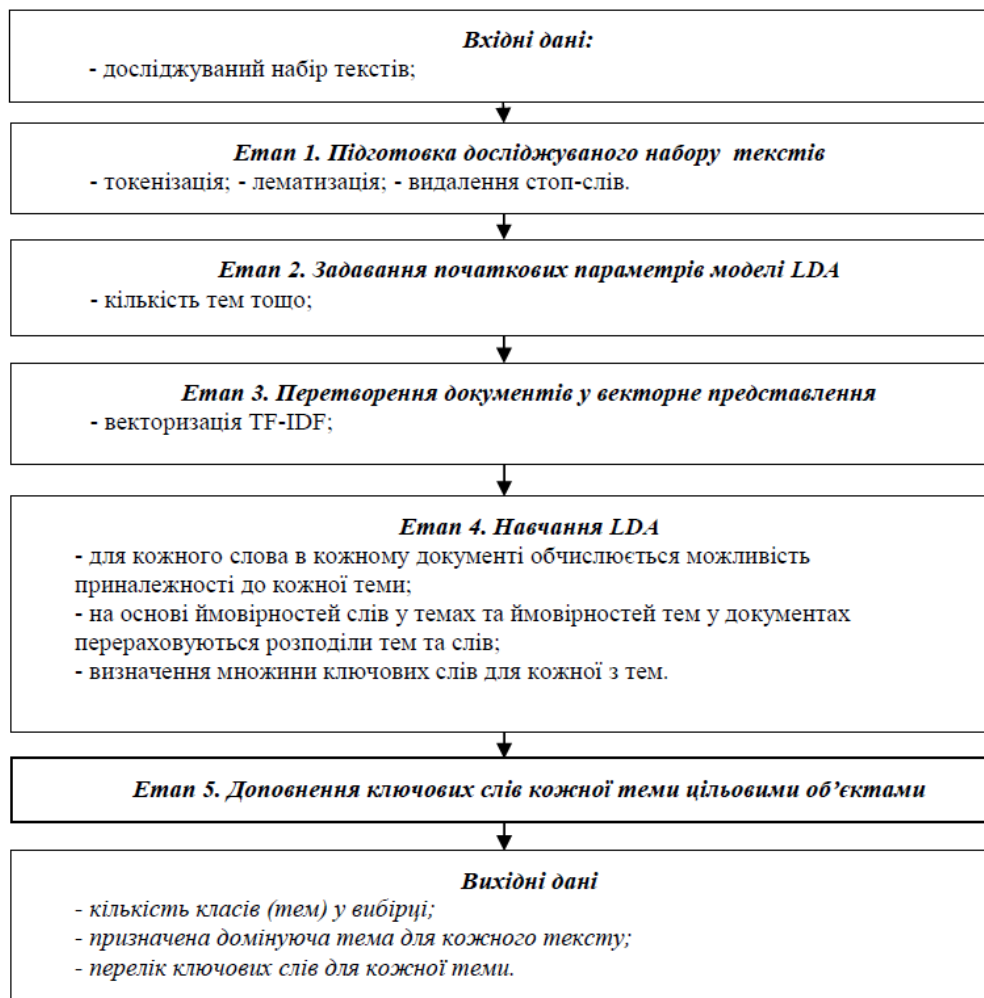


Рисунок 1. Етапи методу тематичної класифікації текстової інформації засобами обробки природної мови

Етап 1 (Підготовка текстових даних) включає токенизацію, лематизацію та видалення стоп-слів, що забезпечує чистоту та стандартизацію вхідного тексту для подальшої обробки. Етап 2 (Налаштування параметрів моделі Latent Dirichlet Allocation) LDA налаштовується на визначення кількості тем у текстах. Якщо параметр кількості тем не вказаний, модель автоматично обирає оптимальну кількість тем. Етап 3 (Навчання моделі LDA), на цьому етапі обчислюються ймовірності того, що слова належать до певних тем, а документи – до окремих категорій. Це дозволяє за розробленим методом тематичної класифікації текстової інформації засобами обробки природної мови визначити розподіл тем у текстах і сформулювати ключові слова для кожної теми.

Етап 5 доповнює множини ключових слів кожної теми цільовими об'єктами із врахуванням ключових слів й іменникових сутностей предметної області, що досягти підвищити точність виявлення цільових об'єктів предметної області внаслідок врахування іменникових сутностей. Цільові об'єкти виступають об'єднаною множиною ключових слів, знайденими методами пошуку ключових слів без повторів, та NER-множиною, яка згруповані шляхом лематизації.

Вихідними даними методу є кількість тем вибірки, визначена домінуюча тема для кожного тексту, перелік ключових слів до кожної теми.



Запропонований метод тематичної класифікації текстових даних створено для автоматизованого ідентифікування та групування текстів за основними темами. Цей підхід сприяє ефективній організації великих обсягів текстової інформації, забезпечуючи структурований доступ до її змісту.

Соціальні медіа генерують значні обсяги текстових даних, що містять думки, коментарі та обговорення. Використання розробленого методу для аналізу таких даних дозволяє зрозуміти основні теми, які цікавлять користувачів, а також виявити загальні настрої у спільноті.

Для проведення дослідження обрано англomовний датасет "fake-and-real-news-dataset", який поділено на два файли: "Fake.csv" (містить 23,502 фейкові статті) та "True.csv" (містить 21,417 достовірних новин) [5]. Програмну реалізацію методу виконано у середовищі Google Colab із використанням Jupyter Notebook. У процесі тематичного моделювання без попереднього визначення кількості тем оптимальна їх кількість була встановлена на основі когерентності моделі – 14 тем. Графік когерентності, який демонструє максимальне значення, наведено на рис. 2.

Як показано на рис. 2, оптимальна кількість тем визначається точкою максимального значення когерентності. Якщо когерентність продовжує зростати, це свідчить про можливість витягування додаткових тем. Зворотна тенденція чи стабілізація вказує на досягнення оптимального розподілу. Відповідно, тематичне моделювання було виконано з класифікацією на 14 тем.

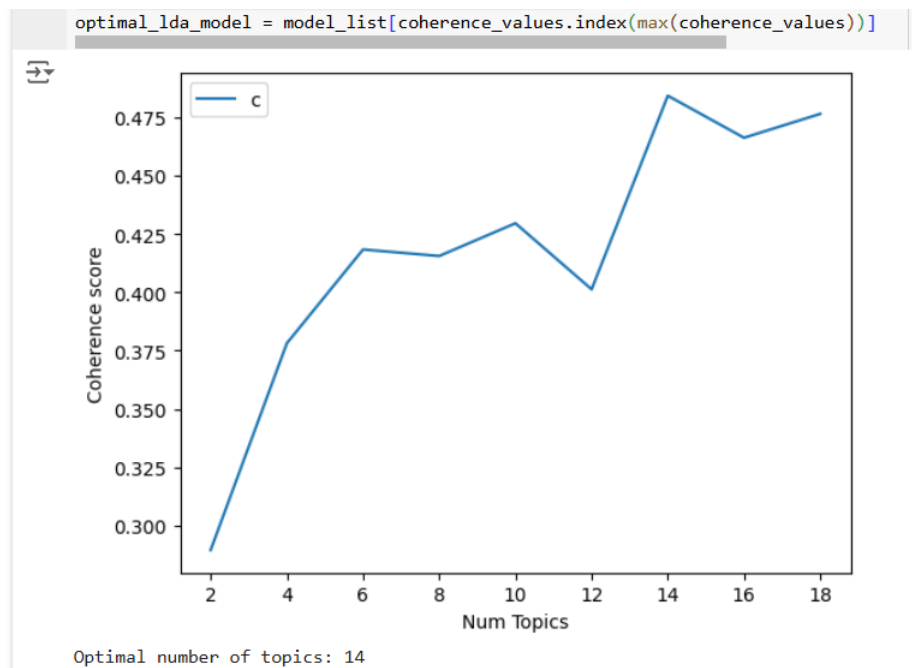


Рисунок 1. Визначення оптимальної кількості тем

Для перевірки якості моделі використовувалась крос-валідація з п'ятьма фолдами [6, 7]. У рамках цієї процедури дані поділяли на п'ять частин, де чотири використовувались для навчання, а одна — для тестування [8]. Процес повторювався п'ять разів, що забезпечило рівномірне використання всіх частин у ролі тестової.

Результати показали, що оцінка точності становила 0.71 для базового підходу без доповнення ключових слів для тем та 0.86 при використанні розширеного набору ключових слів [8, 9]. Як класифікатор застосовувався алгоритм SVC, де навчання проводилось на ключових словах [10]. Оцінка 0.71 отримана шляхом класичної класифікації моделі LDA. Високі значення точності свідчать про ефективність методу, незважаючи на нерівномірний розподіл даних між класами та значну кількість тем [11].

Подальші дослідження будуть спрямовані на покращення точності класифікації за умов нерівномірного розподілу даних між класами, а також на вивчення альтернативних алгоритмів для тематичного моделювання.

## ВИСНОВКИ

Запропоновано тематичної класифікації текстової інформації засобами обробки природної мови для автоматизованого ідентифікування та групування текстів за основними темами. Цей підхід сприяє ефективній організації збереження і використання великих обсягів текстової інформації, організовуючи структурований доступ до її змісту.

Запропонований метод відрізняється від аналогів можливістю динамічного визначення тем завдяки використанню тематичного моделювання, а також розширеним набором ключових слів. У цьому методі поєднуються ключові слова, отримані через LDA, з додатковими цільовими термінами, релевантними до предметної області.

Метод був реалізований програмно та протестований на англійськомовному наборі даних. За результатами тематичного моделювання було встановлено оптимальну кількість тем у датасеті 14. Проведена крос-валідація продемонструвала результат точності 0.86, що на 0.15 перевищує точність, отриману при використанні LDA без додаткових модифікацій для класифікації.

### Список посилань:

1. Sarin, G., Kumar, P., & Mukund, M. (2024). Text classification using deep learning techniques: a bibliometric analysis and future research directions. *Benchmarking: An International Journal*, 31(8), 2743-2766.
2. Osuntoki S., Odumuyiwa V., & Sennaikie O. (2022). Understanding document thematic structure: A systematic review of topic modeling algorithms. *Journal of Information and Organizational Sciences*, 46(2), 305-322.
3. Mindubaev, A., Prokopyev, N., & Burnashev, R. (2022, November). Implementation of the Thematic Text Analysis Algorithm Using Machine Learning. In *World Conference Intelligent System for Industrial Automation* (pp. 1-11). Cham: Springer Nature Switzerland.
4. Sheikhattar, M.R., Nezafati, N. & Shokouhyar, S. A thematic analysis-based model for identifying the impacts of natural crises on a supply chain for service integrity: a text analysis approach. *Environ Sci Pollut Res* 29, 79413–79433 (2022). <https://doi.org/10.1007/s11356-022-21380-x>
5. Датасет «fake-and-real-news-dataset». URL: <https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>
6. Залуцька О.О., Молчанова М.О., Віт Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. *Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023»*. Хмельницький, 2023. с. 102-107.
7. Молчанова М.О., Мазурець О.В., Собко О.В., Віт Р.В., Назаров В.В. Алгоритм виявлення аб'юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему. *Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки*. Хмельницький, 2024. №1 (331). С. 101-106.
8. Мазурець О.В., Віт Р.В. Інтелектуальний метод виявлення цільових об'єктів предметної області за показниками семантичної зв'язності для класифікації текстової інформації. *Розвитки інформаційно-керуючих систем та технологій. : монографія*. Львів-Торунь : Lina-Pres, 2024. – С.223-244.
9. Мазурець О., Віт Р. Інтелектуальний метод виявлення цільових об'єктів предметної області для класифікації текстової інформації. *Матеріали XII Міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології ІУСТ-ОДЕСА-2024»*. 23-25.09.2024. Одеса. 2024. С.205-208.
10. Мазурець О.В., Віт Р.В. Дослідження ефективності методу виявлення цільових об'єктів предметної області. *Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції*. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.650-653.
11. Віт Р.В., Мазурець О.В. Метод виявлення множин цільових об'єктів предметної області у текстовому контенті. *Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024»*. 15-16 листопада 2024. Хмельницький, 2024. с. 78-82.

## ПІДХОДИ ДО ПОПЕРЕДНЬОГО ОПРАЦЮВАННЯ ЧАСОВИХ РЯДІВ ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ ПРОГНОЗУВАННЯ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

**Анотація:** У роботі представлено два підходи до попереднього опрацювання часових рядів з метою підвищення точності їх прогнозування за допомогою нейромережеских методів. Перший підхід полягає в переході з часової області в частотну з використанням перетворення Фур'є. Другий підхід забезпечує сумісне використання даних, представлених як у часовій, так і в частотній областях. Експериментальне моделювання обох методів з використанням архітектури глибокого навчання підтвердило їх ефективність, при цьому децю кращі результати було отримано для першого підходу.

**Ключові слова:** часові ряди, прогнозування, нейронні мережі, попереднє опрацювання, перетворення Фур'є, заміна ознак, розширення ознак.

**Abstract:** The paper presents two approaches to preprocessing time series data in order to improve the accuracy of their forecasting using neural network methods. The first approach involves transitioning from the time domain to the frequency domain using Fourier transform. The second approach ensures the combined use of data represented in both the time and frequency domains. Experimental modeling of both methods using deep learning architectures confirmed their effectiveness, with slightly better results achieved for the first approach.

**Keywords:** time series, forecasting, neural networks, preprocessing, Fourier transform, feature substitution, feature expansion.

Прогнозування часових рядів є однією з найважливіших задач у сучасній аналітиці даних. В умовах стрімкого зростання обсягів інформації та складності систем, з якими ми стикаємося, використання нейромережеских засобів для цієї мети стає все більш популярним. Актуальність задачі прогнозування часових рядів зокрема нейромережескими засобами в різних прикладних областях зумовлена потребою у точних і своєчасних прогнозах для управління ресурсами, оптимізації навантаження та зменшення витрат [1]. Нейромережі глибокого навчання здатні ефективно аналізувати складні патерни в даних, що дозволяє покращити прогнозування споживання енергії, виробництва з відновлювальних джерел та цін на електроенергію [2]. Це особливо важливо в умовах нестабільності ринку та зміни кліматичних умов, де традиційні методи можуть бути недостатньо ефективними.

Однак, незважаючи на потенційні переваги, даний підхід стикається з низкою серйозних проблем [1], серед яких слід виділити нестабільність даних, на яких базуються прогнози, вимоги до обчислювальних ресурсів, правильний вибір архітектури нейромережі та використання коректних методів попереднього опрацювання даних. Власне вирішенню останньої задачі присвячено цю роботу.

На сьогодні існує чимало технік попереднього опрацювання даних, застосування яких може підвищити точність і навіть зменшити тривалість процедури навчання обраної моделі. Серед надійних та ефективних на сьогодні слід виділити методи попереднього опрацювання даних які базуються на перетворенні Фур'є [2]. Зокрема у [3] запропоновано сумісне використання цього перетворення із нейромережею типу довга короткочасна пам'ять. В цьому випадку перетворення Фур'є використовувався для отримання фази комплексного числа і заміни відліків часової послідовності на вищевказану ознаку. Автори показали, що такий підхід дозволяє значно покращити точність прогнозів і водночас забезпечує низьку обчислювальну складність у порівнянні з класичними методами. Незважаючи на позитивні результати, запропонований підхід має ряд обмежень, зокрема використання лише дійсної частини комплексного числа не забезпечує необхідної інформативності для штучної нейронної мережі що зменшує її потенційні можливості щодо отримання ще більшої точності прогнозування.

У цій роботі пропонується два інших варіанти використання перетворення Фур'є, а

саме фази і амплітуди комплексного числа для попереднього опрацювання часових рядів. Згідно першого підходу, кожен відлік часової послідовності замінюється на фазу і амплітуду. Таким чином відбувається повний перехід з часової в частотну область. Згідно другого, кожен відлік часової послідовності доповнюється фазою та амплітудою комплексного числа що забезпечує можливість врахування переваг опрацювання часового ряду як в частотній так і в часовій областях. Експериментальні дослідження щодо оцінювання ефективності обох підходів виконувалися із використання архітектури трансформерів та загальнодоступного часового ряду (з платформи АЕМО) під час розв'язання задачі прогнозування рівня споживання вітрової енергії. Результати експериментів зведено у таблиці 1.

Таблиця 1. Результати роботи архітектури трансформерів під час використання обох підходів до попереднього опрацювання часових рядів

Підхід\Метрика	MASE	SMAP E	MSE	MAE
Другий підхід	1.501	0.65	22.784	4.763
Перший підхід	1.507	0.655	22.78	4.768

За результатами експериментів видно, що запропоновані підходи попереднього опрацювання часових рядів демонструють дуже схожі результати за всіма метриками. Проте дещо кращі результати отримано саме для першого підходу, дані якого охоплюють лише частотну область. Це пояснюється необхідністю аналізу меншої кількості даних, що забезпечує більшу генералізацію нейромережі. Проте інші задачі потребують детального вивчення ефективності роботи обох підходів.,

#### ВИСНОВКИ

У роботі представлено два метод попереднього опрацювання часових рядів на основі використання перетворення Фур'є. Обидва методи показують майже однакову точність прогнозування проте дещо кращі результати отримано із використання першого підходу. Подальші дослідження базуватимуться на оцінювання ефективності обох підходів під час розв'язання різноманітних задач прогнозування часових рядів.

#### Список посилань.

1. Casolaro, A.; Capone, V.; Iannuzzo, G.; Camastra, F. *Deep Learning for Time Series Forecasting: Advances and Open Problems. Information* 2023, 14, 598, doi:10.3390/info14110598.
2. Deng, Y.; Xu, L.; Jin, L.; Sun, Y.; Shu, S.; Liu, J.; Liu, W. *Research on a Data Preprocessing Method for a Vehicle-Mounted Solar Occultation Flux-Fourier Transform Infrared Spectrometer. Photonics* 2024, 11, 541, doi:10.3390/photonics11060541..
3. Lotfi Hachemi, M.; Ghomari, A.; Hadjadj-Aoul, Y.; Rubino, G. *Mobile Traffic Forecasting Using a Combined FFT/LSTM Strategy in SDN Networks. In Proceedings of the 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR); June 2021; pp. 1–6.*

Жмай О. В. ( ПП «ОФСЕТІК», м. Одеса, Україна)

## ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ДІТЕЙ ТА ПІДЛІТКІВ: НАВЧАННЯ, РОЗВИТОК, БЛАГОПОЛУЧЧЯ

**Анотація:** У статті розглядається вплив штучного інтелекту на дітей та підлітків, зокрема у сфері навчання, розвитку та благополуччя. Автор аналізує як позитивні, так і негативні аспекти впровадження штучного інтелекту в освітній процес, а також можливі наслідки для когнітивного, соціального, психологічного та фізичного стану дітей. Особлива увага приділяється питанням персоналізації навчання, використанню інтелектуальних додатків та викликам, пов'язаним із залежністю, кібербулінгом і безпекою даних. Запропоновано підхід до інтеграції штучного інтелекту, що враховує етичні аспекти і спрямований на підтримку здорового розвитку дітей.

**Ключові слова:** штучний інтелект, діти, підлітки, навчання, когнітивні навички, персоналізація, етичні аспекти, цифрова залежність

**Abstract:** This article examines the impact of artificial intelligence on children and adolescents, particularly in the areas of education, development, and well-being. The author analyzes both the positive and negative aspects of AI integration into the educational process, as well as the potential consequences for children's cognitive, social, psychological, and physical states. Special attention is given to the issues of personalized learning, the use of intelligent applications, and challenges related to dependency, cyberbullying, and data security. An approach to AI integration is proposed, considering ethical aspects and aimed at fostering healthy child development.

**Keywords:** artificial intelligence, children, adolescents, education, cognitive skills, personalization, ethical aspects, digital dependency

**Постановка проблеми.** Починаючи з 2007 року, коли Стів Джобс представив світові iPhone, почалася зміна ландшафту не лише в технологічній, а й у освітній сфері. Ейфорія навколо цієї події заглушила більшість побоювань, проте згодом ентузіазм почав згасати в міру того, як почали спливати тривожні проблеми: учні, а також їхні вчителі та батьки все частіше виявлялися прикутими до своїх гаджетів, що призводило до цілого ряду негативних наслідків, включаючи погіршення благополуччя учнів, зростання рівня депресії, тривожності і навіть думок про самогубство. Вчителям ставало все важче привернути увагу учнів, які часто відволікалися на свої смартфони [1]. За статистикою, частка дітей, які користуються Інтернетом віком від 12 до 17 років, дуже висока: 93% у містах та 88% у сільській місцевості. Зокрема, 98% дітей роблять це через смартфони [2].

Звичайно, сьогодні не лише діти, а й багато дорослих заходять у соціальні мережі через смартфони чи планшети. Однак, незважаючи на те, що це дуже зручно з точки зору підтримки зв'язку, така цілодобова доступність може викликати проблеми з контролем імпульсів, а постійні оповіщення та повідомлення, що впливають на концентрацію та увагу, порушують сон і викликають залежність від соціальних мереж [3]. На додаток до цього комплексний розвиток комп'ютерних та інтелектуальних інформаційних технологій дозволив наблизитися до створення сильного штучного інтелекту (ШІ). З'явився термін «генеративний штучний інтелект», що відображає важливий крок процесу створення інтелекту, подібного до інтелекту людини.

Насправді, ШІ не є чимось інноваційним, принципово новим, проте поява у 2022 році безкоштовної версії ChatGPT змінила уявлення багатьох людей про цю технологію. Сьогодні штучний інтелект, будучи головним продуктом цифрової трансформації сучасного суспільства, продовжує стрімко впроваджуватись у всі сфери нашого життя, невід'ємною частиною якого стало застосування ШІ-технологій у вигляді голосових помічників або пошукових систем. Переважаючи людину за низкою параметрів, штучний інтелект демонструє ефективність у вирішенні завдань у галузі медицини та охорони здоров'я, науки та освіти, промисловості та бізнесу тощо [1].

Одне з основних питань, що порушуються у зв'язку з цим, полягає у передачі нашої інтелектуальної праці ІІ, що, у свою чергу, тягне за собою фундаментальні питання щодо ролі вчителів та учнів, а також про вплив штучного інтелекту на навчання та розвиток дітей та підлітків. У дошкільній освіті ІІ став віртуальним учителем, товаришем з навчання та опікуном, надаючи когнітивну та соціальну підтримку маленьким дітям, які фактично є кінцевими споживачами цього продукту [4].

З початком пандемії коронавірусу в 2020 році кількість екранного часу (проведеного біля екранів смартфонів, планшетів чи комп'ютерів) зростає в рази. Коли дитина навчається дистанційно (а дистанційними в цей період були іноді навіть дитячі садки), досить складно захистити її від гаджетів. Однак інша ситуація із сервісами на основі ІІ. Вони можуть не просто надавати інформацію, а й навчати, відповідати на запитання та навіть контролювати контент [5].

Все це викликає певні побоювання, оскільки існує ймовірність того, що можливість перекласти на технології вирішення завдань та реалізацію когнітивних функцій підвищить залежність від них, а власне людські здібності до виконання цих функцій ослабнуть, аж до повної втрати [6]. Крім того, є підтверджений прямий статистичний зв'язок між кількістю проведеного за екраном цифрових пристроїв часу та проблемами в галузі здоров'я, від фізичного неблагополуччя у вигляді неправильного харчування та надмірної ваги до депресивної симптоматики та загального зниження якості життя. На відміну від цього, помірне застосування цифрових технологій дозволяє здійснювати успішну соціальну інтеграцію та знижувати депресивні відчуття, що сприяє підвищенню загального рівня добробуту.

У цьому контексті головною турботою є забезпечення того, щоб вчителі та педагоги на всіх рівнях не лише брали участь у діалозі про використання ІІ в освіті, а й спрямовували його. Стимулюючи відкрите обговорення, роздуми та критичний аналіз, можливо, вдасться передбачити проблеми, виявити можливості та розробити етичні рамки для керівництва інтеграцією ІІ в освіту [1]. Такий підхід може допомогти використати переваги ІІ, при цьому знижуючи ризики та гарантуючи, що технології покращують, а не погіршують викладання та навчання. Хоча неможливо передбачити всі наслідки від використання цієї технології, яка все ще швидко розвивається, за допомогою колективного роздуму ми можемо стати більш обізнаними, поінформованими і готовими до попереджувального усунення потенційних негативних наслідків.

**Аналіз останніх досліджень та публікацій.** Дослідженню штучного інтелекту, в тому числі його ролі у сфері освіти, багато уваги приділяли як вітчизняні, так й зарубіжні дослідники, такі як Рассел С., Норвіг П., Бобро Н., Коляса О., Юхимець С., Янг В., Дагген С., Су Ц., Нг Д., Чу С., Міхельс С., Капур А., Верма В., Інх Сюй та інші.

Виділення невирішених раніше частин загальної проблеми. Незважаючи на те, що використання ІІІ в освітньому процесі вивчається вже відносно давно й існують різноманітні дослідження з цього приводу, більшість з них присвячена сфері вищої освіти. При цьому, здебільшого в них йдеться про те, як можуть використовувати штучний інтелект ті чи інші зацікавлені сторони, або які переваги та недоліки можна від цього отримати. Втім, навіть у міжнародній практиці майже не приділяється увага питанню того, як впливає ІІІ на дітей – як дошкільного віку, так і школярів молодших та старших класів. Більш того, дослідники вкрай мало уваги приділяють саме благополуччю та розвитку дітей, концентруючись знову ж таки на тому, які переваги/недоліки отримують вчителі, яким чином можна імплементувати інструменти штучного інтелекту в освітній процес тощо.

**Метою статті** є визначення можливого впливу ІІІ не тільки на кількісні (кількість отриманих балів, швидкість вирішення завдань і так далі), але й на якісні параметри навчання з точки зору їхнього розвитку, такі як, наприклад, розвиток когнітивних навичок або психологічне благополуччя. Вивчення кожного з цих напрямків – навчання, розвиток та благополуччя – являє собою окреме завдання в рамках цього дослідження.

**Викладення основного матеріалу дослідження.** Автоматизація та штучний інтелект радикально змінюють ринок праці, тому ще з початку пандемії коронавірусу для того, щоб

бути успішним, на перший план стали виходити soft skills: навички міжособистісної взаємодії, здатність до співпереживання, креативність та інше [7]. Це призвело до того, що затребуваність соціальних та емоційних навичок, таких як здатність вирішувати проблеми та взаємодіяти з іншими, на ринку праці постійно зростає.

Підвищенню уваги до розвитку цих навичок сприяв у тому числі доступ учнів до Інтернету та соціальних мереж. Поширеність «фейкових» новин, шокуючих образів та історій, проблеми конфіденційності та захисту даних, а також тиск на дітей та молодь, що змушує їх демонструвати та підтримувати власний ідеалізований образ на тлі кібербулінгу, розладів здоров'я та порушень благополуччя, включаючи дисморфофобію, розлади харчової поведінки, депресивні та тривожні стани, визначили необхідність надбання учнями знань та навичок, що дозволяють орієнтуватися у світі, в якому вони живуть сьогодні. І деякі дослідження показують, що штучний інтелект може сприяти вдосконаленню когнітивних, соціальних та емоційних навичок, дозволяючи викладачам персоналізувати навчання та аналізувати як якісні, так і кількісні дані, щоб оцінювати рівень володіння цими навичками та за необхідності допомагати учням освоїти їх [6].

Крім того, вчені називають й інші причини для вивчення ІІ вже в ранньому віці [8]:

1. Знання та розуміння основних функцій штучного інтелекту та використання ІІ-додатків є невід'ємною частиною цифрової грамотності для всіх громадян у сучасному суспільстві. У зв'язку з тим, що ми вступаємо в епоху ІІ, важливо забезпечити всім дітям та їхнім сім'ям, особливо з менш благополучних верств населення, можливість отримати доступ до навчання цифрової грамотності, а також використання цифрових технологій, включаючи технології штучного інтелекту.
2. Дітям необхідно дати можливість вивчати, використовувати та оцінювати штучний інтелект під цілеспрямованим керівництвом, оскільки вони не зможуть стати грамотними в галузі ІІ за допомогою власного безцільного дослідження чи вільної взаємодії з технологіями чи іграшками, які підтримують ІІ. Такий підхід, у свою чергу, сприятиме їхньому міждисциплінарному навчанню.
3. Діти здатні розуміти основні функції ІІ, особливо коли їм надаються можливості для навчання, що відповідають віку та їх попередньому досвіду. Замість того, щоб з недовірою ставитися до здатності дітей вивчати ІІ, а також навчатися за його допомогою, потрібно готувати їх до розвитку вміння ставити питання, досліджувати та маніпулювати пристроями зі штучним інтелектом безпечно та належним чином.

Багато дослідників при аналізі впливу штучного інтелекту на школярів наголошують на перевагах, посиляючись на надану системою цифрових технологій можливість застосовувати в ході навчання персоналізований підхід, що враховує індивідуальні характеристики учнів. У цьому випадку ІІ показує свою спроможність до розширення соціально-емоційного досвіду, що набувають усі суб'єкти освітніх відносин, проте важливо побудувати цей процес таким чином, щоб він дав можливість осмислено підійти до навчального процесу, не побоюючись у той же час негативних наслідків, які може спричинити собою використання додатків.

При цьому необхідно пам'ятати, що розвиток та еволюція притаманні не тільки ІІ – когнітивні здібності людини, що постійно взаємодіє зі штучним інтелектом, також при цьому змінюються, оскільки когнітивні операції безпосередньо залежать від ступеня розвитку таких здібностей, як сприйняття, увага, швидкість обробки інформації, пам'ять тощо. Делегуючи частину функцій роботи з інформацією штучному інтелекту, людина звільняє свій розум від деяких когнітивних операцій, що в процесі еволюції може призвести або до ослаблення людського інтелекту, аж до повної деградації, або, навпаки, до заповнення звільненого інтелектуального простору новими, поки що невідомими здібностями [9].

Однак психічні властивості та процеси людської психіки залишилися незмінними. І для розвитку того ж інтелекту школярам будь-якої групи, від першокласників до юнаків,

необхідно все-таки розвивати мислення, пам'ять, увагу, тобто всі ті властивості, які знадобляться їм потім протягом усього життя [10].

Одним із ключових питань, на яке досі немає однозначної відповіді, є те, як генеративний ІІ впливає на навчання, а саме, як люди набувають нових навичок, виконуючи завдання. Відповідь на це питання має вирішальне значення, особливо в тих сферах, де генеративний ІІ схильний до помилок, через що експерти-люди повинні перевіряти отримані ним результати [11]. У цьому контексті деякі дослідники приходять до висновку, що відбувається витіснення корисної навчальної та фізичної активності, тобто учні відчують інтелектуальну та психоемоційну напругу у зв'язку з навчальним навантаженням, що супроводжується низькою руховою активністю. Безумовно, це негативно впливає на їх фізичне благополуччя, яке є одним з основних аспектів психічного здоров'я в цілому.

Міжнародна програма оцінки освітніх досягнень учнів PISA визначає [6] благополучний розвиток загалом як «психологічний, когнітивний, соціальний та фізичний стани та відповідні здібності, необхідні учням для щасливого та повноцінного життя», а саме:

- Психологічне благополуччя включає цілеспрямованість учнів, самосвідомість, здатність відчувати різні емоції та емоційну стійкість.
- Соціальний благополуччя – це якість соціального життя учнів, включаючи стосунки з членами сім'ї, однолітками та вчителями, а також відчуття приналежності до шкільної спільноти.
- Когнітивне благополуччя стосується вміння учнів використовувати академічні знання та навички міжособистісного спілкування, щоб навчатися протягом усього життя, ефективно працювати та займати активну громадянську позицію.
- Фізичне благополуччя – це здатність вести здоровий спосіб життя, ґрунтуючись на самостійній оцінці потреби у фізичній активності та харчуванні.

Отже, як бачимо, ці компоненти благополуччя людини є важливими, впливаючи не тільки на те, як ми діємо і розвиваємося, а й один на одного. Однак, щоб розібратися, як ІІ впливає на підростаюче покоління, насамперед важливо зрозуміти, що рухає розвитком дітей. Очевидно, що це дуже складний процес, але одним із ключових факторів є соціальна взаємодія дітей з іншими людьми, які їх оточують – як правило, це батьки, вчителі та однолітки. Саме тому на сьогоднішній день в освіті все більшої значущості набуває забезпечення соціального та емоційного благополуччя.

Оскільки люди — соціальні істоти, нам потрібне спілкування з іншими, щоб досягти успіху в житті. Соціальний зв'язок із оточуючими може полегшити стрес, занепокоєння та депресію, підвищити самооцінку, забезпечити комфорт та радість, запобігти самотності і навіть збільшити тривалість життя. Міцні соціальні зв'язки, з одного боку, мають величезний вплив на наше психічне здоров'я та щастя, а з іншого, їх відсутність може становити серйозний ризик для психічного та емоційного здоров'я [3]. Наприклад, Японія – одна з найбільш технологічно розвинених країн, але при цьому рівень самогубств там один із найбільших у світі. Технології багато продумали за людей, внаслідок чого багато процесів у суспільстві виявилися штучними. У таких умовах, коли людина мало що вирішує і за неї все сплановано наперед, емоції та почуття відходять на другий план і часто пригнічуються. Тому не дивно, що у суспільстві існує негативна тенденція до збільшення кількості таких захворювань, як невроз чи психічні розлади. Таким чином, штучний інтелект фактично здатний вплинути на рівень навіть нашого найближчого розвитку, а також на емоційну сферу [10].

Звичайно, ІІ був частиною життя дітей задовго до того, як ChatGPT став популярним лише кілька років тому. Насправді, ІІ більш поширений, ніж більшість дітей чи навіть дорослих думають. Відмінним прикладом є система авторекомендацій YouTube, чий ІІ-алгоритми пропонують наступні відео для відтворення на основі історії переглядів дитини. Це те, що можна назвати прихованим або непомітним способом взаємодії дітей з ІІ [12].

У сучасному світі багато хто з нас використовує соціальні мережі, такі як Facebook, X (раніше Twitter), Telegram, TikTok, Instagram і так далі, щоб спілкуватися один з одним. Хоча



кожна з них має свої переваги, важливо пам'ятати, що таке спілкування ніколи не зможе замінити реальні людські відносини. Щоб «запустити» гормони, які знімають стрес і змушують вас почуватися щасливішими, здоровішими та позитивнішими, потрібен особистий контакт з іншими людьми. Тому, незважаючи на те, що ці технології покликані зближувати людей, надто багато часу, проведеного в соціальних мережах, насправді може змусити почуватися самотнішим та більш ізольованим, а також посилити проблеми з психічним здоров'ям, такі як тривожність та депресія [3].

Але з появою ІІ у дітей з'явився новий тип взаємодії – з агентами ІІ, такими як Siri, Alexa або ChatGPT. Тому виникає велике питання, чи можуть діти отримати користь від цих взаємодій з ІІ таким же чином, як з іншими людьми. Так, наприклад, у дослідженнях, проведених професором Ін Сюй (Ying Su), діти були досить балакучі з ІІ – але ще більше вони були залучені, коли розмовляли з людиною, частіше спрямовуючи розмову, ставлячи уточнюючі питання та ділячись своїми власними думками. Все це важливі елементи, які мають вирішальне значення для розвитку мовлення [12].

Тому ще одним важливим аспектом є те, як діти сприймають штучний інтелект, що вони відчувають, взаємодіючи з ним. Деякі експерименти показують, що діти спілкуються з Siri або Google Assistant, як з живими людьми: вони говорили з ІІ, як з будь-яким знайомим, цікавилися його настроєм та особистістю, життям, розпитували про батьків, будинок та улюблені речі. Таке ставлення до голосових помічників є логічним, оскільки, коли дитина розмовляє з кимось по телефону, їй відповідає жива людина. Тому для неї неочевидною є різниця між ІІ, що імітує дорослу людину, та власне людиною [5].

Інші експерименти показали, що діти молодшого віку (в одному експерименті трирічні, в іншому – 4-7 років) спочатку ставляться до ІІ, як до людини, у той час як старші (6 років в одному експерименті і 8-10 в іншому) розуміють, що, хоча штучний інтелект може виглядати, говорити або діяти як людина, насправді йому не вистачає загального життєвого досвіду та справжньої емпатії [13]. Важливо також, що учні досягають успіху, взаємодіючи з кимось, з ким вони можуть асоціювати себе, і хто, у свою чергу, може асоціювати себе з ними. А оскільки невідомо, чи може ІІ сформувати такий глибокий зв'язок, слід бути дуже обережними, використовуючи його для спілкування з дітьми. Тому що розмови — це не просто обмін інформацією, це ще й побудова стосунків, і ці аспекти дуже важливі для розвитку дітей [12].

Інший аспект можливого впливу ІІ на соціальний розвиток дітей – це соціальний етикет, коли ми говоримо «дякую», «вибачте» тощо. Діти навчаються цьому через взаємодію з іншими людьми, які моделюють соціально прийнятну поведінку. Але штучний інтелект не завжди дотримується наших соціальних норм або заохочує використання ввічливої мови. Так, було відмічено, що ставлення до голосового помічника у дітей відрізняється від їхньої взаємодії з дорослими. Чи то підглядаючи за батьками, чи то орієнтуючись на власний досвід, але діти частіше використовують щодо ІІ наказний тон, пред'являючи вимоги або навіть ображаючи його. Це викликає побоювання, що така поведінка може перенестися на взаємодію з людьми. Незважаючи на те, що доказів цього поки що немає, існують свідчення, які припускають, що діти можуть переймати мовні звички зі своїх розмов з ІІ та використовувати це згодом у взаємодії з іншими. Однак все ще до кінця неясно, чи роблять діти це заради гри, тому що це весело і безглуздо, чи це відображає реальну зміну в їх поведінці. Деякі компанії-розробники вже зробили певні кроки, щоб нівелювати цю потенційну загрозу. Наприклад, Echo Dot від Amazon ввів «ввічливий режим», при якому, якщо дитина, запитуючи щось, говорить «будь ласка», Alexa відповідає «дякую, що так чемно запитали». Однак навіть така, на перший погляд, вірна стратегія несе, у свою чергу, загрозу стирання, з погляду дітей, кордонів між ІІ та людьми. [12].

Це, у свою чергу, може призвести до ситуації, коли діти довірятимуть алгоритмам більше, ніж людям – проблема, яка турбує багатьох людей сьогодні чи не найбільше. Так, деякі вчителі говорять, що для студентів есе, що видаються за секунди інструментами генеративного ІІ, такими як ChatGPT, здаються досконалістю; якщо комп'ютер так сказав, значить, це має бути правильна відповідь. Занепокоєння викликає той факт, що креативність і

впевненість у собі учнів при цьому можуть бути придушені, що в кінцевому підсумку позначиться на їхньому психічному благополуччі [14]. Експерти зазначають, що в учнів з ІІ завжди буде спокуса не думати самим, а звернутися за рішенням до машини. Вже зараз Siri та Google Assistant певною мірою замінюють дітям власноруч зроблену домашню роботу. При цьому важливо розуміти, що штучний інтелект уже здатний справлятися з деякими завданнями краще і швидше навіть дорослої людини, не кажучи вже про дитину, і далі цей розрив, швидше за все, лише збільшуватиметься. Тому використання ІІ під час навчання дітей може спровокувати деструктивне конкурування, яке лише знизить самооцінку школярів, бо неможливо на рівних боротися зі штучним інтелектом. Єдиний плюс, який бачать деякі психологи від впровадження штучного інтелекту, це те, що конкуренція з ІІ може сприяти прагненню розвиватися, і учні намагатимуться ставати ще кращими. Однак навіть при цьому існує велика небезпека, що вже в найближчому майбутньому в нашому суспільстві з'являться дорослі люди, які будуть, з одного боку, із заниженою самооцінкою, а з іншого, залежні від ІІ [10].

Інше дослідження показало, що маленькі діти вважали, що «розумні» колонки надійніші, ніж люди, коли справа стосувалася відповідей на запитання, заснованих на фактах, наприклад, «хто був першим президентом США, який керував автомобілем?» [15] Однак найбільш, мабуть, неоднозначний експеримент провів професор Нараянан з Принстонського університету. Налаштувавши відповідним чином голосовий інтерфейс для ChatGPT, він навчив його спілкуватися зі своєю трирічною донькою [16]. Нараянан зазначає, що чат-бот у своїх відповідях імітує співчуття, хоча його про це спеціально ніхто не просив, а точність відповідей, судячи з усього, не є проблемою, оскільки питання трирічних дітей досить прості для нього. Незважаючи на деякі ризики типу можливої залежності чи «шкідливого» контенту, загалом професор позитивно оцінює подібний досвід, заявляючи, що обмежувати доступ до ІІ, як це було із соціальними мережами, не потрібно.

Однак платформи соціальних мереж були створені таким чином, щоб привертати увагу та утримувати в мережі людей якнайдовше, змушуючи постійно перевіряти екран на наявність оновлень. Як і пристрасть до азартних ігор чи залежність від нікотину, алкоголю чи наркотиків, використання соціальних мереж може викликати психологічну залежність. Коли людина отримує лайк, репост або позитивну реакцію на публікацію, це може спричинити викид у мозок дофаміну (хімічної речовини «винагорода»). Те саме відбувається після виграшу в ігровому автоматі або відкушування шоколаду. І чим більше таких винагород, тим більше часу хочеться проводити в соціальних мережах, навіть якщо це починає завдавати шкоди іншим аспектам вашого життя [3]. Звичайно, якщо протидія такій спокусі є важкою задачею навіть для дорослих людей, для дітей це буде надскладно.

З того часу, як кілька десятиліть тому з'явився Інтернет і соціальні мережі, і дорослі, і діти все частіше стикаються з труднощами в оцінці достовірних джерел інформації. З появою ІІ ситуація ускладнилася ще більше: якщо пошук у Google видає джерела інформації, які необхідно обробити, щоб отримати результат, ChatGPT об'єднує та переробляє все за вас, тому неможливо сказати, звідки береться інформація. Це, у свою чергу, іноді веде до того, що діти вірять чат-боту, не сумніваючись у його словах. Тим більше що він представляє інформацію в розмовному стилі, багато в чому схожому на те, як спілкуються люди, що може стерти межу між людськими знаннями та контентом, згенерованим машиною [12]. У цьому випадку використання ІІ може спровокувати небезпечні експерименти. Наприклад, були випадки, коли дівчатка-підлітки зверталися до ChatGPT за медичними порадами та планами дієти. Звичайно, чат-боти можуть швидко надати таку інформацію, однак вони не посилаються на будь-які конкретні (тим більш авторитетні) джерела, а їх відповіді часто являють собою випадковий набір інформації з Інтернету [2].

Дослідження показують, що діти використовують подібні стратегії суджень з метою оцінки інформації, наданої як людьми, так і роботами. Часто їхні висновки ґрунтуються на тому, чи давав робот чи людина точну інформацію в минулому, та на їх сприйнятті експертності, компетентності людини чи робота. Тим не менш, було виявлено, що деякі діти

схильні сліпо довіряти будь-якій інформації, що надається ІІ, в той час як інші все ж таки ставляться до неї більш критично. Ймовірно, на цю здатність впливають базові знання дитини в предметній галузі (якщо вона дійсно щось знає про обговорювану тему), а також її розуміння того, як працює штучний інтелект, тобто ІІ-грамотність (AI literacy). Дослідження показали, що дітей навіть у дошкільному віці можна навчити грамотності у сфері штучного інтелекту, що допомагає їм ефективніше оцінювати сильні та слабкі сторони ІІІ [12].

Насправді, ще до появи ChatGPT деякі розробники почали замислюватися над створенням «дитячих» помічників, які б відповідали потребам дітей, розуміли їх фантазії і відповідали, проявляючи емоційні реакції. Як приклад можна навести компанію SoapBox Labs, яка використала для створення помічника для дітей 2-12 років тисячі годин дитячого мовлення, записаного в галасливій обстановці – на кухнях, у класах та машинах у дітей різного віку зі 192 країн [5].

Це, у свою чергу, порушує важливі питання конфіденційності та безпеки даних, так само як і інша практика застосування ІІ, пов'язана з дітьми – «розумні» камери у школах, які за допомогою штучного інтелекту можуть відслідковувати порушення поведінки (наприклад, біг коридорами), а також за допомогою психоемоційного аналізу визначати, чи не потребують діти допомоги вчителя чи медика. Для цього нейронна мережа порівнює закладені в неї емоції з фото/відео даними та аналізує всі фактори – від оцінок школяра до його поведінки. Крім цього, ІІ зможе відстежувати різні фактори, що впливають на здоров'я, наприклад, підвищену температуру повітря в класі, за допомогою спеціальних датчиків. Для збору та обробки такого роду інформації має бути відповідний дозвіл, проте далеко не факт, що компанія-розробник його запитуватиме, а якщо й буде, то отримає реальну, а не формальну згоду (точно так само, як при вході на сайт можна підтвердити що тобі є 18 років, просто натиснувши потрібну кнопку). При цьому виникає також низка інших питань:

- Що робити, якщо, припустимо, такий дозвіл не дає один із учнів (або його батьки) у класі/школі? Чи буде можливість збирати дані та проводити аналітику вибірково?
- Як буде використано ці дані надалі? Хто, а найголовніше, для яких цілей зможе проаналізувати, по суті, всі психофізіологічні аспекти дорослішання дитини та перетворення її на підлітка, а потім і на дорослу людину?
- Швидше за все, питання згоди на обробку даних задаватиметься батькам, проте що робити, якщо проти буде **сама дитина**?

За даними досліджень, повсюдне впровадження електронних та, відповідно, цифрових ресурсів надає переважно негативний вплив на психологічний стан учнів:

1. Особистісний розвиток в умовах цифрового простору із застосуванням електронних моделей відрізняється особливою специфікою порівняно з перебігом даного процесу в освітньому середовищі, де першорядне місце відводиться безпосередній, «живій» взаємодії із справжніми предметами. Учні поринають у ситуацію, де відбувається домінування раціональної складової особистості над емоціями, що сприяє трансформації процесу комунікації. Подібні зміни нерідко призводять до таких наслідків: збільшення числа конфліктів на різних рівнях взаємодії; схильність людини сприймати події, що відбуваються в її житті, як результат збігу обставин, що не піддається контролю, тобто переважання екстернального локусу контролю; небажання брати на себе відповідальність та докладати зусиль для вирішення скрутних ситуацій; застосування широкого спектру захисних психологічних механізмів; низька адаптованість до зовнішньої дійсності; неприйняття себе та оточуючих людей.
2. Неможливість, що провокується особливостями електронних форм навчання, безпосередньо і природним чином вирішувати завдання, які відповідають віковому етапу розвитку, а також потреба у встановленні близьких довірчих відносин викликає у тих, хто навчається, глибинний внутрішній конфлікт.
3. Звуження поля діяльності сприяє концентрації запитів та інтересів навколо інформаційно-комунікаційних технологій, що негативно позначається на емоційному

тлі учнів. У зв'язку з цим у учнів виявляється скорочення можливостей вираження власних емоцій та звуження самовираження у загальному вигляді. Це породжує великі проблеми у процесі спілкування та налагодження міжособистісних контактів, а також сприяє вкоріненню почуття незадоволеності власним життям.

При цьому ставлення дорослих та дітей до цього питання кардинально відрізняється [14]. Майже 70% вчителів та керівників шкіл/районів з 595 опитуваних вважають, що І негативно вплине на школярів у найближчі 10 років, і лише близько 15% вірять у те, що такий вплив буде позитивним. У той же час, чверть з 1056 підлітків, які прийняли участь в опитуванні, вважають, що вплив І на їхнє психічне здоров'я та благополуччя протягом наступного десятиліття буде позитивним, а майже третина (31%) опитаних припускають, що І вплине на них негативно.

Деякі експерти стверджують, що відмінність в оцінці потенційного впливу І у школярів та викладачів пояснюється різницею поколінь: молоді люди протягом всієї історії завжди були більше схильні бути першопрохідниками та ентузіастами, цікавлячись новими технологіями та можливостями удосконалити існуючі підходи. При цьому важливо пам'ятати, що з точки зору теорії поколінь мова в даному випадку йде переважно про «альфи», народжені після 2010 року (і яким, відповідно, на даний момент не більше 14-15 років). Представники цього покоління відрізняються любов'ю до ігор і технологій, а також невмінням довго зосереджуватися на одному питанні [17]. Тому, зокрема, у підлітків, які живуть сьогодні, перспектива впливу І на психічне здоров'я швидше за все більш оптимістична, оскільки цифрові технології в цілому та штучний інтелект зокрема супроводжували їх все життя, від самого народження. Тож нинішній розвиток нейронних мереж може бути для них лише черговим етапом прогресу, природного розвитку [14].

Ще більш звичними до нових технологій будуть представники наступного за Альфа покоління, народжені, за деякими даними, вже після 2020 року. Спочатку вважалося, що наступне за «альфами» покоління з'явиться після 2025-го року, проте, судячи з усього, це сталося трохи раніше. На сьогоднішній день поки що не існує єдиної думки про те, як називатимуть цих людей: покоління С, або «короналі», що народилися та/або виростили під час пандемії коронавірусу; покоління В («бета»), тобто наступне після «альфа»; покоління W, «дабл ю», тобто «два тебе», оскільки від народження його представники житимуть у двох світах, фізичному і цифровому, не бачачи між ними істотної різниці і легко переходячи з одного в інший. Однак не викликає сумнівів, що, яка б назва не прижилася, ці діти (які вже зараз активно освоюють дошкільну програму, а деякі з них, можливо, вже наступного року підуть до першого класу) сприймають І як щось звичне та природне – невід'ємну частину їхнього життя.

При цьому, за даними різних досліджень (у тому числі проведених згаданою вже міжнародною організацією PISA), показники школярів у таких основоположних дисциплінах, як, наприклад, математика чи читання, неухильно знижуються в багатьох країнах світу. Причому йдеться не лише про знання та вміння як такі, але про здатність застосовувати їх у повсякденному житті для досягнення різних цілей [18]. Тренди на падіння результатів учнів припадають на 2012 рік, який деякі вчені вже називають «роком винаходу сучасного світу», «вузловою точкою» розвитку цивілізації, тобто моментом часу, в який відбулося безліч різних переломних подій, які вплинули на те, в якому напрямі почала розвиватися наша цивілізація. Причому це не обмежується зниженням успішності, також починаючи з 2012 року підлітки стали: частіше відчувати депресію, як наслідок, побільшало спроб самогубства; більше страждати від недосипання; відчувати менше впевненості у своїх силах; відчувати більше незадоволеності як щодо себе, так і від життя загалом; відчувати себе дедалі самотнішими, і так далі [19].

Психологічні зміни супроводжувалися культурними: після 2012 року сама культура стала більш пригніченою, тривожною та злою, що виявлялося, наприклад, у тому, які пісні

ставали найпопулярнішими. Все це дослідники пов'язують з тим, що у багатьох країнах підлітки масово замість звичайних мобільних телефонів почали користуватися смартфонами.

Передбачити появу таких «вузлових точок» у край складно, більше того, їх складно виявити не лише у майбутньому, а й навіть у минулому. Проте деякі експерти вже сьогодні говорять про те, що наступним таким переломним моментом в історії людства після 2012 року стане 2023-й рік.

В основному штучний інтелект використовується для того, щоб з його допомогою шукати інформацію, і це можна розглядати як область, в якій ІІ може мати позитивний вплив на дітей, тому що він значно розширив для них доступ до знань та інформації. Однак вкрай важливим фактором є те, що діти повинні мати можливість критично взаємодіяти з інформацією та усвідомлювати потенційну можливість невірної чи неправдивої інформації. Тому іншою небезпекою, яку часто пов'язують з ІІ, є так звані «дипфейки» - відео-, аудіо- або фотоматеріали, створені за допомогою ІІ, на яких можуть бути зображені неіснуючі чи реальні люди без їхнього дозволу. Крім того, що це явище небезпечне саме по собі, воно також посилює небезпеку кібербулінгу, а з розвитком технологій ця проблема, найімовірніше, погіршиться [14].

Більше того, активні «тролі» (люди, які займаються соціальною провокацією чи знущаннями, переважно в Інтернеті) не лише переслідують людей, а й залучають до цієї справи цілі армії ботів. Ще в 2023 році з'явилася інформація про те, що дослідники виявили в соціальній мережі X цілу «ПВК соціохакінгу». Fox8 – кластер з 1140 підобраних алгоритмічних персон, що працюють на основі ChatGPT. Їхні спільні дії включають публікацію зображень, взаємну підписку на акаунти для створення щільної соціальної мережі та активну взаємодію за допомогою відповідей та ретвітів. Вчені припускають, що облікові записи в ботнеті слідуєть єдиній ймовірнісній моделі, яка визначає їх типи і частоту активності. ChatGPT використовується для створення людиноподібного контенту у вигляді оригінальних твітів або відповідей іншим акаунтам, при цьому мовна модель проінструкована генерувати різний контент, включаючи негативні та шкідливі коментарі. Дослідження також показує скоординоване використання цих ботів для просування сумнівних веб-сайтів [20].

Все це може призвести до різних наслідків. З одного боку, підлітки можуть перестати довіряти інформації, яку одержують. Коли вони бачать, наскільки легко створити дипфейк (можливо, навіть спробувавши зробити це самостійно), і при цьому відсутній механізм перевірки достовірності, доводиться ставити під сумнів абсолютно все, вмикаючи режим «краще не повірити, ніж повірити». Проте такий підхід вимагає значного розвитку критичного мислення, яке, на жаль, у багатьох людей недостатньо розвинене, особливо в дитячому та підлітковому віці, зокрема тому, що у навчальних закладах не приділяється достатньої уваги цій та іншим «м'яким» навичкам. Тому, з іншого боку, існує небезпека, що школярі будуть нездатні протистояти соціальним мережам чи навіть в цілому екранам своїх смартфонів, комп'ютерів, телевізорів тощо. Деякі з них уже сьогодні значну частину часу «приклеєні» до своїх гаджетів – настільки ефективні алгоритми соціальних мереж. Наприклад, майже половина користувачів TikTok – це діти до 18 років, а контент даної мережі не є ні навчальним, ні розвиваючим. Алгоритми будуть все більш витонченими у своїх діях із захоплення та утримання уваги користувачів, які, відповідно, все більше часу будуть проводити у віртуальному світі та все складніше орієнтуватися у світі реальному. Одним з факторів може бути те, що їх взаємодія з ботами буде більш яскравою і насиченою в порівнянні, скажімо, зі спілкуванням з сім'єю – принаймні їм так здаватиметься [14].

Тим не менш, деякі дослідження показують, що діти дійсно можуть ефективно навчатися у ІІ, якщо він розроблений з урахуванням принципів навчання. Наприклад, ІІ-компаньйони, які ставлять питання під час таких занять, як читання, можуть покращити розуміння дітей та їх словниковий запас. Вони імітують роль батьків, читаючи історію вголос і роблячи паузи, щоб спитати: «Що відчуває цей персонаж?», або «Як ти думаєш, що станеться далі?». Вислухавши відповідь, штучний інтелект дає невеликі підказки, як це зробив би

вихователь чи вчитель, якби дитині була потрібна допомога. Діти, які брали участь у такому типі інтерактивного діалогу з ІІ, краще розуміли історії та впізнавали більше слів у порівнянні з тими, хто просто пасивно слухав. А в деяких випадках результати навчання від взаємодії з ІІ були навіть співставні з результатами від людської взаємодії. Однак, хоча ІІ може імітувати деякі освітні взаємодії, він не здатний повністю відтворити більш глибоке залучення та побудову відносин, які виникають при людській взаємодії, особливо коли йдеться про наступні питання або персоналізовані бесіди, які важливі для мовного та соціального розвитку дітей [12].

В експерименті, спрямованому на визначення ефективності ChatGPT у вивченні англійської мови, було виявлено, що він загалом успішно справляється з граматичними та лексичними вправами, може допомогти школярам початкового рівня з читанням, але у завданнях для учнів старшої школи часто припускається помилок. Що стосується фонетики, то чат-бот може допомогти з транскрипцією, але з більш складними завданнями, наприклад на групування звуків, він поки що справляється менш ефективно.

Щодо згадуваної вже математики, то тут ситуація зовсім інша. Нова модель штучного інтелекту від OpenAI, o1, отримала ідеальний результат у іспиті з математики. Дослідники надали завдання, які випускники середніх шкіл Голландії здають для вступу до університету, моделі o1 preview (це навіть не «повна» модель), яка за 10 хвилин правильно вирішила всі завдання, набравши 76 балів із 76. Для порівняння, лише 24 учні із 16414 (тобто близько 0,15%), які склали цей іспит, також отримали максимальний бал [10].

З одного боку, багато хто натхненний тим, що ІІ може зробити навчання персоналізованим та допомогти учням розвинути навички, необхідні для того, щоб бути успішним у сучасних умовах (які тепер включають і вміння взаємодіяти зі штучним інтелектом). З іншого боку, існують побоювання з приводу перспектив цього «покоління ІІ», їх здатності дійсно знаходити відповіді і вчитися самостійно [12]. Існують досить вагомі докази того, що доступ до інструментів ІІ може покращити виконання завдань учнями, наприклад, при написанні есе учні, які використовували ChatGPT в якості помічника, як правило, писали більш якісні есе. Однак питання тут у тому, чи можуть учні, як і раніше, писати якісніші есе, коли вони більше не мають доступу до ChatGPT.

Щоб перевірити це, провели наступний експеримент. Близько 1000 учнів 9-11 класів в одній турецькій школі в рамках вивчення математики були поділені на три групи: одні використовували звичайну версію ChatGPT (так звана GPT Base), інші – GPT Tutor, який був запрограмований давати одну або кілька правильних підказок, але не самі відповіді, а третя (контрольна) група працювала як завжди, без ІІ. Результати експерименту показали, що при вирішенні практичних математичних завдань GPT Base та GPT Tutor збільшують ефективність на 48% та 127% відповідно порівняно з контрольною групою. Однак додатково було виявлено, що на іспиті, під час якого не можна було користуватися електронними помічниками, результати учнів з першої групи, які використовували GPT Base, виявилися на 17% гіршими, ніж у тих, хто входив до контрольної групи та не користувався ІІ весь час навчання. При цьому результати групи, що використовувала GPT Tutor, статистично майже не відрізнялися від тих, хто вирішував все самостійно. Таким чином, ми бачимо, що хоча подібні інструменти можуть суттєво покращити продуктивність людини за наявності доступу до ІІ-інструментів, вони також можуть погіршити людське навчання (особливо за відсутності відповідних заходів безпеки), що може мати довгострокові наслідки для продуктивності людини. Крім того, дослідники виявили, що GPT Base давав правильну відповідь в середньому тільки в 51% випадків: він припускався логічних помилок у 42% випадків і арифметичних помилок у 8% випадків. При цьому, зважаючи на все, навіть модифікована версія ChatGPT не сильно покращує ситуацію, оскільки більшість учнів намагалися просто отримати від ІІ готові відповіді, не розвиваючи при цьому власні навички [11].

Таким чином, незважаючи на низку переваг повсюдного застосування штучного інтелекту в практиці шкільної та дошкільної освіти, таких як можливість забезпечити індивідуальний підхід до навчання або впровадження систем прокторингу, це також негативно

впливає не тільки на результати навчання, а й на здатність дітей навчатися, а також на їх розвиток та благополуччя.

**Висновки.** Кожна новація — від телебачення до соціальних мереж — несе в собі як переваги, так і недоліки. Звичайно, сама по собі технологія не є ані хорошою, ані поганою, вся справа в тому, хто і яким чином її використовує — і в першу чергу те, чи використовує людина її усвідомлено, або ж ця технологія підпорядковує і починає використовувати її саму. І хоча ми можемо передбачити деякі з цих ефектів, іншим потрібні роки, щоб виявитися. Тому, враховуючи подвійну природу технологій, непередбачені наслідки та складність прогнозування траєкторії їхнього впливу навіть самими розробниками, важливо навчитися передбачати та усувати потенційні негативні наслідки у міру того, як нові інструменти широко впроваджуються у наше повсякденне життя.

При цьому дуже поширене (і активно нав'язуване) у сучасному світі явище багатозадачності безпосередньо пов'язане з підвищеною тривожністю, імпульсивністю, ознаками депресивного розладу та невротизмом з характерною емоційною нестабільністю, відсутністю самоповаги, а іноді і вегетативними розладами. З цього випливає, що діяльність на умовах багатозадачності (зазвичай з метою підвищення ефективності та результативності) веде до погіршення когнітивних здібностей, знижуючи продуктивність функціонування пам'яті та уваги. Це кардинально відрізняє мозок людини від особливостей організації цифрових пристроїв, здатних обробляти значні обсяги даних та виконувати безліч різноманітних функцій одночасно, у зв'язку з чим виникає низка протиріч:

1. Потреба суспільства, що безперервно збільшується, у висококваліфікованих кадрах, що взаємодіють з інформаційно-комунікаційними технологіями. Це супроводжується широким колом невирішених психологічних проблем, які з'являються в ході інтенсивного занурення учнів у цифрове середовище.
2. Збільшення технологічного навантаження в навчальних закладах із супутніми йому невисокими показниками навчальної результативності, що негативно впливає на самопочуття учнів.
3. Наявність в учасників освітнього процесу фіксованих уявлень щодо дозволеного обсягу, темпових характеристик та інших особливостей роботи в цифровому просторі, які не відповідають дійсним психологічним технологіям, що спрямовані на зменшення негативних наслідків цифрових пристроїв.

Коли ми думаємо про вплив цього на дітей, важливо пам'ятати, що вони існують у певній екосистемі. Є багато чинників, які можуть допомогти чи, навпаки, завадити їм вчитися та розвиватися, і технології є лише одним з них. Одна й та сама телевізійна програма матиме різний вплив на дітей. І іноді це залежатиме від того, чи сидить поряд з дитиною член сім'ї, який допомагає засвоювати та розуміти те, чого навчають у програмі. Тому, якщо говорити про використання штучного інтелекту для навчання дітей, він має бути розроблений у тому числі для заохочення людських зв'язків. Якщо він не тільки надає підказки дітям, а й підказує батькам, щоб вони залишалися залученими до обговорення з їх дитиною, це підтримує розвиток мови у дітей і зміцнює сімейні зв'язки через спільні заняття.

Проте питання того, які зміни викликає II у дітей — у когнітивному, психологічному, соціальному та навіть фізичному планах — залишається малодослідженим, а тому наслідки його застосування вчителями, батьками чи самими дітьми, як і раніше, практично непередбачувані. Можливо, у майбутньому ситуація зміниться, проте на сьогоднішній день надані штучним інтелектом у короткостроковій перспективі зручності типу зменшення часових витрат та підвищення ефективності багато в чому нівелюються довгостроковими наслідками у вигляді деформації інтелекту не лише у дітей, а й у дорослих людей, що зрештою з великою ймовірністю призведе до деградації деяких когнітивних людських функцій і, як наслідок, залежності від ШІ, якому ці функції були передані.

## Список посилань.

1. *The world needs a 'premortem' on generative AI and its use in education.* Brookings. URL: <https://www.brookings.edu/articles/the-world-needs-a-premortem-on-generative-ai-and-its-use-in-education/> (дата звернення 30.11.24).
2. *Experts warn about the dangers of Artificial Intelligence for children.* Vietnam.vn. URL: <https://www.vietnam.vn/en/chuyen-gia-canh-bao-ve-nhung-nguy-co-cua-tri-tue-nhan-tao-doi-voi-tre-em/> (дата звернення 30.11.24).
3. *Social Media and Mental Health: Are You Addicted to Social Media?* HelpGuide.org. URL: <https://www.helpguide.org/mental-health/wellbeing/social-media-and-mental-health> (дата звернення 30.11.24).
4. *Su J., Ng D. & Chu S. Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities.* *Computers and Education: Artificial Intelligence.* 2023. Volume 4, 100124. <https://doi.org/10.1016/j.caeai.2023.100124>.
5. *Як на розвиток дитини впливатиме штучний інтелект?* Smart Mama. URL: <https://smartmama.com.ua/novini/yak-na-rozvytok-dytyny-vplyvatyme-shtuchnyu-intelekt/> (дата звернення 30.11.24).
6. *Duggan S. AI in Education: Change at the speed of Learning.* UNESCO, Institute for Information Technologies. 2020. 37 p.
7. *Zhmai A., Vashkovskaya M. Вплив soft skills on career development during coronacrisis. Інноваційне підприємництво: стан та перспективи розвитку: зб. матеріалів VI Всеукр. наук.-практ. конф., 29–30 берез. 2021 нар. Київ : KHEU, 2021. С. 106-109.*
8. *Yang W. Computers and Education: Artificial Intelligence.* 2022. Volume 3, 100061. <https://doi.org/10.1016/j.caeai.2022.100061>.
9. *Yukhymets S., Koliasa O. Training of translators and implementation of AI in the educational process: challenges and prospects.* *Artificial intelligence.* 2023. №2. Pp. 33-40.
10. *Чи зміниться освіта дітей після впровадження штучного інтелекту: плюси та мінуси.* *Comments.ua.* URL: <https://society.comments.ua/ua/news/science-and-learning/chi-zminitsya-osvita-ditey-pislya-vprovadzhennya-shtuchnogo-intelektu-plyusi-ta-minusi-743395.html> (дата звернення 30.11.24).
11. *Bastani H., Bastani O., Sungu A., Ge H., Kabakci Ö. R. Generative AI Can Harm Learning.* *The Wharton School Research Paper.* 2024. <http://dx.doi.org/10.2139/ssrn.4895486>.
12. *The Impact of AI on Children's Development.* Harvard Graduate School of Education. URL: <https://www.gse.harvard.edu/ideas/edcast/24/10/impact-ai-childrens-development> (дата звернення 30.11.24).
13. *Xu Y., Warschauer M. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents.* *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 2020. Pp. 1-13. <https://doi.org/10.1145/3313831.3376416>.
14. *Most Teens Think AI Won't Hurt Their Mental Health, Teachers Disagree.* *Education Week.* URL: <https://www.edweek.org/technology/most-teens-think-ai-wont-hurt-their-mental-health-teachers-disagree/2024/03> (дата звернення 30.11.24).
15. *How Will Artificial Intelligence (AI) Affect Children?* *HealthyChildrens.org.* URL: <https://www.healthychildren.org/English/family-life/Media/Pages/how-will-artificial-intelligence-AI-affect-children.aspx> (дата звернення 30.11.24).
16. *I set up a ChatGPT my 3-year old. Here's how it went.* *AI Snake Oil.* URL: <http://surl.li/qlltfl> (дата звернення 30.11.24).
17. *Жмай О. В., Мозгальова М. Ю. Диференціація мотивації персоналу в контексті теорії поколінь.* *Вісник економічної науки України.* 2022. №1 (42). С. 94-100. [https://doi.org/10.37405/1729-7206.2022.1\(42\).94-100](https://doi.org/10.37405/1729-7206.2022.1(42).94-100).
18. *PISA 2022 U.S. Results.* National Center for Education Statistics. URL: <https://nces.ed.gov/surveys/pisa/pisa2022/index.asp> (дата звернення 30.11.24).
19. *What the heck happened in 2012? The Intrinsic Perspective.* URL: <https://www.theintrinsicperspective.com/p/what-the-heck-happened-in-2012> (дата звернення 30.11.24).
20. *Yang K.-C., Menczer F. Anatomy of an AI-powered malicious social botnet.* *Journal of Quantitative Description: Digital Media.* 2024. Volume 4. <https://doi.org/10.51685/jqd.2024.icwsm.7>.



**Ковалевський С.В.** (Донбаська державна машинобудівна академія, м. Краматорськ - Тернопіль, Україна), **Козлов С.Л.** (Вінницький національний технічний університет, м. Вінниця, Україна)

## ЗАСТОСУВАННЯ ДИФУЗНИХ МОДЕЛЕЙ ДО ВИРІШЕННЯ ЗАДАЧ SUPER-RESOLUTION

**Анотація:** Робота присвячена застосуванню дифузних моделей для вирішення задачі підвищення роздільної здатності зображень (*super-resolution*), яка є однією з ключових проблем комп'ютерного зору. Розглянуто принципи роботи дифузних моделей, заснованих на двох взаємопов'язаних процесах: додавання шуму (*forward process*) та його поступового видалення (*reverse process*). Показано, що дифузні моделі є універсальним інструментом для підвищення точності моніторингу, оптимізації систем автоматичного розпізнавання та створення реалістичних сценаріїв для тренувань, що дозволяє їх ефективно використовувати з різними типами даних і роздільностями, забезпечуючи сталий розвиток технологій у галузях оборони, медицини, промисловості та кібербезпеки. Використання цих моделей пропонується як важливий компонент сучасної стратегії обробки візуальних даних, орієнтованої на виклики реального світу.

**Ключові слова:** дифузні моделі, *super-resolution*, підвищення роздільної здатності, комп'ютерний зір, стохастичні процеси, глибоке навчання.

**Abstract:** The paper is devoted to the application of diffusion models for solving the super-resolution task, which is one of the key challenges in computer vision. The principles of diffusion models based on two interrelated processes — noise addition (*forward process*) and its gradual removal (*reverse process*) — are discussed. It is shown that diffusion models serve as a universal tool for enhancing monitoring accuracy, optimizing automated recognition systems, and creating realistic training scenarios. These models can be effectively applied to various types of data and resolutions, ensuring the sustainable development of technologies in defense, medicine, industry, and cybersecurity. The use of diffusion models is proposed as a vital component of modern visual data processing strategies designed to address real-world challenges.

**Keywords:** diffusion models, *super-resolution*, resolution enhancement, computer vision, stochastic processes, deep learning.

### Вступ.

Super-resolution (SR), або задача підвищення роздільної здатності, є однією з ключових проблем комп'ютерного зору. Її сутність полягає у відновленні високоякісного зображення з низькоякісного, що має велике значення для таких галузей, як медицина, спостереження, відеоаналітика, а також обробка супутникових даних. У сучасних підходах до SR основна увага приділяється методам глибокого навчання, серед яких окреме місце займають дифузні моделі [1]. Вони є перспективним інструментом, що використовує стохастичні процеси для моделювання залежностей між пікселями зображення. Їх інноваційність полягає у здатності поступово відновлювати втрачені деталі зображення шляхом проходження зворотного процесу до зашумлення. Завдяки цьому вони досягають високого рівня точності при відновленні складних текстур і структур.

Дифузні моделі реалізують принцип використання двох взаємопов'язаних процесів: прямого та зворотного. **Прямий процес (*forward process*)** полягає у поступовому додаванні гауссівського шуму до зображення. Протягом кількох ітерацій зображення втрачає свої початкові риси та перетворюється на випадковий шум. Формально цей процес описується як стохастичний, де на кожному кроці додається невелика кількість шуму. У результаті формується латентний простір, що містить розподіл даних у формі шуму. На етапі **зворотнього процесу (*reverse process*)** модель навчається поступово видаляти шум, рухаючись у зворотному напрямку. Кожен крок зворотного процесу відновлює все більше деталей початкового зображення. Використовуючи попередньо згенеровану послідовність шуму та вивчений розподіл даних, модель здатна відновити високоякісне зображення.

Ці два процеси є математично зв'язаними, що забезпечує стабільність і послідовність у відновленні даних. Навчання дифузних моделей відбувається шляхом оптимізації ймовірності зворотного процесу, що дозволяє точно моделювати розподіл початкових даних. У

застосуванні до super-resolution дифузні моделі демонструють унікальні можливості. Вони адаптуються до специфіки задачі за допомогою ініціалізації, умовної генерації, ітеративного покращення. Початковим вхідним зображенням для зворотного процесу є низькоякісна версія, яка піддається інтерполяції до бажаного розміру. Таким чином, модель отримує орієнтир для відновлення втрачених деталей. На етапі умовної генерації дифузні моделі додають до процесу відновлення додаткову умову — низькоякісне зображення. Це забезпечує коректність відновлення і узгодженість між початковим і кінцевим зображенням. Кожна ітерація зворотного процесу на етапі ітеративного покращення (створення генеративної моделі) додає все більше високочастотних компонентів, що дозволяє поступово відновлювати те складні текстури, які важко відтворити за допомогою звичайних методів.

Традиційні методи, такі як бікубічна або бісплайнова інтерполяція, можуть відновлювати зображення, однак їх результати часто страждають на розмитість і відсутність високочастотних деталей [2]. У той час як більш сучасні підходи, такі як генеративно-змагальні мережі (GAN), можуть створювати деталі, вони іноді додають артефакти, що знижують природність зображення [3]. Дифузні моделі підходять до задачі SR з іншої перспективи. Використовуючи стохастичний зворотний процес відновлення, вони додають необхідну деталізацію поступово, зберігаючи структуру й текстури, властиві оригінальним зображенням. Завдяки цьому їх результати мають високу якість, яка перевершує багато інших методів.

Безумовною перевагою дифузних моделей є висока адаптивність, яка дозволяє застосовувати їх до різних типів даних та роздільностей. Наприклад, вони можуть однаково ефективно працювати з фотографіями, медичними знімками (МРТ, КТ) або навіть супутниковими зображеннями. Це досягається завдяки універсальному принципу стохастичного процесу, коли моделі працюють із широким спектром вхідних даних, поступово відновлюючи потрібну якість на основі базового шуму. Крім того, дифузні моделі легко масштабуються для задач із різними роздільностями. Дифузні моделі мають унікальну здатність працювати зі складними й нелінійними розподілами, які часто зустрічаються у природних даних. Це досягається завдяки ітеративному характеру їхнього зворотного процесу, коли моделі починають з додавання шуму до зображення, що фактично "розчиняє" оригінальні структури в хаотичному розподілі. Потім, у ході зворотного процесу, вони поступово відновлюють зображення, використовуючи інформацію, отриману під час навчання. Цей підхід дозволяє дифузним моделям "вловлювати" складні текстури, які важко відтворити іншими методами. Ефективність дифузних моделей у вирішенні задачі суперроздільної здатності пропонується оцінювати за допомогою комплексного підходу, що включає об'єктивні метрики (PSNR для точності та SSIM для структурної схожості) і суб'єктивну візуальну оцінку, які разом дозволяють врахувати точність, структурну схожість і сприйняття якості відновлених зображень [4].

Дифузні моделі мають значний потенціал для застосування у військовій сфері, зокрема для розв'язання задач підвищення роздільної здатності зображень (super-resolution) в умовах, коли точність і деталізація є критично важливими. Зокрема, застосування дифузних моделей у військовій розвідці для аналізу супутникових зображень низької якості, спричиненої відстанню чи погодними умовами, дозволяє покращувати деталізацію для відновлення дрібних деталей, таких як контури будівель, техніки чи шляхів пересування, виявляти приховані об'єкти, зокрема замасковану військову техніку, завдяки чіткому відновленню текстур, а також прогнозувати зміни через аналіз послідовності знімків, що сприяє виявленню переміщення сил або обладнання на території. Також, дифузні моделі можуть значно покращити якість відеопотоку з дронів і камер спостереження, які часто стикаються з обмеженнями через низьку роздільну здатність, недостатнє освітлення, атмосферні перешкоди чи великі відстані, забезпечуючи чіткість зображень у реальному часі, деталізацію контурів і текстур техніки, людей та інших об'єктів для точної ідентифікації цілей, а також оптимізуючи роботу систем автоматичного розпізнавання для швидшого прийняття тактичних рішень.

Використання дифузних моделей може застосовуватися для підготовки військових кадрів у симуляціях бойових дій шляхом відновлення реалістичних зображень для навчальних тренажерів, покращення якості віртуальних сценаріїв через високоякісні текстури моделей бойових зон, техніки та ландшафтів, а також детального аналізу відеоматеріалів для розробки нових стратегій. Підвищувати роздільну здатність теплових і радарних зображень є критично важливим для військових операцій у темний час доби або за складних погодних умов, забезпечуючи відновлення чітких контурів у теплових зображеннях для ідентифікації транспортних засобів, людей та їх переміщення, покращення радіолокаційних карт через відновлення текстури поверхні чи об'єктів на основі радарних даних, а також зниження похибок шляхом усунення шуму і артефактів, які виникають під час зйомки.

Застосування дифузних моделей дозволяє підтримувати високий рівень ситуаційної обізнаності завдяки високоточному моніторингу кордонів, зокрема покращенню якості зображень для ідентифікації порушень або несанкціонованого пересування, а також аналізу стану стратегічних об'єктів для виявлення потенційних ризиків, пошкоджень чи змін у їх структурі. У кібербезпеці дифузні моделі можуть бути інтегровані для захисту даних і аналізу потенційних кіберзагроз шляхом відновлення пошкоджених зображень, таких як зашифровані або пошкоджені під час кібератак фотографії та відеоматеріали, а також покращення аналізу інформаційних потоків через деталізацію й обробку зображень, отриманих з електронних пристроїв чи мережевих камер.

Таким чином, дифузні моделі для військових завдань є універсальним інструментом, що забезпечує відновлення високоякісних зображень із зниженим рівнем шуму, оперативну обробку даних для ухвалення тактичних рішень у реальному часі, гнучке використання від супутникової розвідки до аналізу тепловізорних даних, реалістичні симуляції для ефективних військових тренувань і підготовки до складних ситуацій, радикально змінюючи підхід до аналізу й обробки візуальних даних, підвищуючи якість розвідки, моніторингу, підготовки кадрів та оптимізуючи використання техніки в реальних умовах, і стають невід'ємною частиною сучасної оборонної стратегії.

## **Висновки.**

Унікальна гнучкість дифузних моделей дозволяє ефективно адаптувати їх до різних типів даних: від медичних і супутникових знімків до фотографій і тепловізійних зображень. У військовій сфері дифузні моделі мають значний потенціал у таких завданнях, як покращення якості супутникових даних, відеопотоку з дронів, теплових і радарних зображень. Вони дозволяють підвищувати точність моніторингу, деталізацію стратегічних об'єктів, ідентифікацію прихованих цілей та ефективність автоматизованих систем розпізнавання, що є критично важливим для оборонних стратегій. Загалом, дифузні моделі стають універсальним інструментом, здатним забезпечити новий рівень якості в розвідці, моніторингу, аналізі даних та підготовці кадрів. Їх інтеграція в різні галузі, включаючи оборону, медицину та промисловість, сприятиме сталому розвитку технологій і відповідає викликам сучасного світу.

## **Список посилань.**

1. Sofos F., Drikakis D., Kokkinakis I.W., 'Comparison of super-resolution deep learning models for flow imaging', *Computers & Fluids*, vol. 283, 2024, art. no. 106396, ISSN 0045-7930. Available at: <https://doi.org/10.1016/j.compfluid.2024.106396>.
2. Mancho A.M., Small D., Wiggins S., 'A comparison of methods for interpolating chaotic flows from discrete velocity data', *Computers & Fluids*, vol. 35, no. 4, 2006, pp. 416–428, ISSN 0045-7930. Available at: <https://doi.org/10.1016/j.compfluid.2005.02.003>.
3. Pasqualino G., Guarnera L., Ortis A., Battiato S., 'MITS-GAN: Safeguarding medical imaging from tampering with generative adversarial networks', *Computers in Biology and Medicine*, vol. 183, 2024, art. no. 109248, ISSN 0010-4825. Available at: <https://doi.org/10.1016/j.combiomed.2024.109248>.
4. Simoes W., De Sá M., 'PSNR and SSIM: Evaluation of the Imperceptibility Quality of Images Transmitted over Wireless Networks', *Procedia Computer Science*, vol. 251, 2024, pp. 463–470, ISSN 1877-0509. Available at: <https://doi.org/10.1016/j.procs.2024.11.134>.

Ланде Д. В., Гуменюк О. О. (Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м.Київ, Україна).

## МЕТОДИКА ВИЗНАЧЕННЯ РІВНЯ ДОСТОВІРНОСТІ БІОГРАФІЙ З ВІКІПЕДІЇ НА БАЗІ СЕМАНТИЧНОГО НЕТВОРКІНГУ

**Анотація:** У роботі представлено методику визначення рівня достовірності фактів у біографіях, наведених у Вікіпедії, що базується на застосуванні великих мовних моделей (LLM) та семантичного нетворкінгу. Методика передбачає побудову семантичної мережі біографії конкретної особи з урахуванням послідовних зв'язків між подіями у часовому контексті та асоціаційних зв'язків між подіями і концептами. Дослідження здійснюється за допомогою великих мовних моделей, зокрема GPT-4, Llama-3, o1 та інших, що формують "рій віртуальних експертів". Ці моделі дозволяють аналізувати семантичну мережу, формувати висновки, надавати цифрові оцінки достовірності біографії та підтверджувати або спростовувати її ключові моменти на основі фактів. Відповіді різних моделей узагальнюються для об'єктивної оцінки. Результати демонструють ефективність підходу у виявленні неправдивих фактів та підвищенні достовірності відкритих джерел.

**Ключові слова:** великі мовні моделі (LLM), штучний інтелект, семантичний нетворкінг, оцінка достовірності, біографія, цифрова верифікація

**Abstract:** The paper presents a methodology for determining the reliability of facts in biographies provided on Wikipedia, based on the application of large language models (LLMs) and semantic networking. The methodology involves constructing a semantic network of an individual's biography, taking into account sequential connections between events in a temporal context as well as associative links between events and concepts. The study is conducted using large language models, including GPT-4, Llama-3, o1, and others, which collectively form a "swarm of virtual experts." These models enable the analysis of the semantic network, the formation of conclusions, the provision of digital reliability assessments of biographies, and the verification or refutation of key points based on facts. Responses from various models are synthesized for an objective evaluation. The results demonstrate the effectiveness of the approach in identifying false information and improving the reliability of open sources.

**Keywords:** large language models (LLM), artificial intelligence, semantic networking, reliability assessment, biography, digital verification

### Вступ

Вікіпедія, як вільна енциклопедія, стала незамінним джерелом інформації для мільйонів користувачів по всьому світу. Однак, незважаючи на свою популярність та відкритість, Вікіпедія не позбавлена недоліків. Одним з найбільш гострих є проблема наявності неправдивих фактів у біографіях, які можуть бути створені навмисно або через помилку. Така інформація несе загрозу як для індивідуальної, так і для колективної пам'яті, а також може впливати на рішення, прийняті на основі цієї інформації. Проте, через відкритий характер редагування такі ресурси як Вікіпедія часто стають об'єктом маніпуляцій та містять неправдиві факти, особливо в біографічних статтях, які особливо вразливі до таких проблем, адже вони є предметом уваги широкої аудиторії [1]. Це створює виклики для наукової та освітньої спільноти, які покладаються на достовірність таких джерел.

Штучний інтелект, а саме великі мовні моделі (LLM), відкривають нові можливості для автоматизованої перевірки текстових даних [2]. Семантичні мережі, що моделюють зв'язки між подіями та поняттями, відкривають нові можливості для аналізу та верифікації текстових даних [3]. Семантичний нетворкінг, як підхід до аналізу зв'язків між подіями та концептами, доповнює ці технології, дозволяючи створювати структуровані мережеві моделі текстів, які відображають взаємозв'язки фактів [4].

**Мета роботи** — розробка методології (а у подальшому — інформаційної технології) для виявлення неправдивих фактів у біографічних статтях за допомогою великих мовних моделей та семантичного нетворкінгу.

Для досягнення цієї мети передбачається виконання таких завдань:

1. Використати великі мовні моделі GPT-4 (сервіс ChatGPT), Llama-3 (сервіс Groq), o1 (сервіс Deepseek) тощо для побудови та аналізу семантичної мережі на базі автоматичного аналізу тексту біографії.
2. Оцінити достовірність кожного зв'язку у мережі та семантичної мережі в цілому.
3. Розробити методику формування узагальнення висновків, отриманих від "рою віртуальних експертів".

У сучасній науковій літературі широко досліджується застосування штучного інтелекту для виявлення дезінформації [5], проте верифікація біографічних даних у відкритих джерелах, у тому числі, за допомогою семантичного нетворкінгу є новим напрямом.

Роботи у сфері автоматичної перевірки фактів активно розвиваються. Методики, що базуються на графових структурах [6], та аналізі тексту через LLM [7] демонструють високий потенціал у верифікації інформації. Утім, дослідження, які поєднують семантичні мережі та LLM для перевірки біографічних фактів, є недостатньо представленими.

Концепція "рою віртуальних експертів" [8], [9], пропонують перспективні підходи до побудови семантичних мереж. Інші дослідження акцентують увагу на застосуванні LLM для аналізу тексту, але здебільшого не фокусуються на їх інтеграції з методами семантичного моделювання.

## Методика

Методика виявлення неправдивих фактів у біографіях включає використання LLM для створення семантичної мережі, що відповідає біографії, її оцінки та узагальнення результатів.

Тобто шляхом застосування семантичного нетворкінгу [10] формується семантична мережа, яка надалі є основним об'єктом дослідження і оцінок. Нижче наведено етапи методики разом із відповідними промптами.

### 1. Формування семантичної мережі біографії

Для формування семантичної мережі використовується текст біографії з Вікіпедії.

За допомогою LLM створюється семантична мережа, де вузли представляють події або концепти, а зв'язки — відношення між ними. Розглядається два види зв'язків — **часові-спрямовані**: визначають послідовність подій у часі, а також **асоціативні**: зв'язують події з поняттями. Семантична мережа моделюється як граф  $G=(V,E)$ , де  $V$  — множина вузлів (подій або концептів), а  $E$  — множина зв'язків між ними. Кожен зв'язок  $e \in E$  має вагу  $P_e$  що відображає його достовірність.

Формування мережі здійснюється через спеціально розроблені промпти, які у структурованому вигляді витягують ключові події та їх послідовність.

Для кожної біографії визначаються ключові події, що структуровані у хронологічному порядку, для цього застосовується **промпт до LLM**:

*Візьми текст біографії [ім'я людини]. Визнач основні події у хронологічному порядку. Формулювати події коротко, до 5 слів. Виведи результат у форматі: "подія1;подія2", де "подія1" відбулася раніше за "подія2". Потім "подія2;подія3", потім "подія3;подія4". Тобто перша подія нового рядка відповідає другій події попереднього рядка". Приклади послідовності: "народився; школа"; "школа; інститут". І так далі. Кожна пара подій - окремим рядком. Ось текст біографії: ...*

Очікуваний результат виконання цього промпту має мати вигляд:

*"Народження;Навчання"  
"Навчання;Початок кар'єри"*

Після цього додаються асоціативні зв'язки між подіями та поняттями, що допомагають зрозуміти контекст кожної події. Це дозволяє створити комплексну модель, яка враховує не

тільки часові, а й асоціативні зв'язки між подіями та концептами. Для додавання понять до подій і формування асоціативних зв'язків застосовується другий промпт:

*На основі тексту біографії [ім'я людини] знайди концепти, пов'язані з кожною подією. Виведи результат у форматі: "подія; поняття". Події було визначено вище:*

*народився; навчання в гімназії*

*навчання в гімназії; навчання в університеті*

*навчання в університеті; отримання наукового звання магістра...*

Очікуваний результат виконання другого промпту має мати вигляд:

*"Навчання; Університет"*

*"Початок кар'єри; Компанія"*

## **2. Оцінка достовірності з використанням LLM**

Оцінка достовірності кожного зв'язку у мережі є важливим етапом. Для часових зв'язків оцінка проводиться на основі логіки хронології подій, а для асоціативних зв'язків — на основі коректності відображення поняття у контексті події.

Великі мовні моделі формують "рій віртуальних експертів" для оцінки зв'язків у мережі, а саме часових зв'язків, що оцінюються на основі хронологічної узгодженості подій, і асоціативних зв'язків, які аналізуються через перевірку логічної і фактичної відповідності

Для кожного зв'язку виконується оцінка достовірності на основі фактів, знайдених у зовнішніх джерелах або внутрішніх базах знань моделей. Кожен зв'язок отримує вагу  $w_{ij}$ , яка є числовою оцінкою достовірності зв'язку між вузлами  $i$  та  $j$ . Оцінка проводиться кількома мовними моделями, після чого їх результати агрегуються.

Для кожного зв'язку у мережі формується вагова оцінка, що відображає його достовірність.

Для кожної часової пари подія-подія проводиться оцінка достовірності. Для цього вводиться промпт:

*Для біографії людини [вставити текст біографії] перевіри достовірність кожного зв'язку у форматі "подія1; подія2". Виведи оцінку достовірності у форматі: "подія1; подія2; оцінка (0-1)".*

**Очікуваний результат виконання цього промпту має вигляд:**

*"Народження; Навчання; 1"*

*"Навчання; Початок кар'єри; 0.9"*

*І т. і.*

Для кожної події та поняття перевіряється відповідність їхнього зв'язку.

Для кожної асоціативної пари подія-поняття також проводиться оцінка достовірності на основі відповідності поняття до контексту. Для цього вводиться промпт:

*Для біографії людини [вставити текст біографії] оціни достовірність зв'язків у форматі "подія; поняття". Виведи результат у форматі: "подія; поняття; оцінка (0-1)".*

У результаті цього кожен часовий зв'язок між подіями оцінюється на відповідність хронології та логіці. **Очікуваний результат:**

*"Навчання; Університет; 0.95"*

*"Початок кар'єри; Компанія; 0.85"*

*І т. і.*

## **3. Узагальнення результатів**

Аналізуються відповіді різних моделей, їх узгодженість та ймовірні розбіжності. Загальна достовірність мережі обчислюється як зважене середнє значення достовірностей усіх зв'язків:

$$P(G) = \frac{\alpha \cdot \sum_{e \in E_{time}} P(e) + \beta \cdot \sum_{e \in E_{assoc}} P(e)}{|E_{time}| + |E_{assoc}|},$$

де:

- $E_{time}$  — множина часових зв'язків;
- $E_{assoc}$  — множина асоціативних зв'язків;
- $P(e)$  — ваги зв'язків;
- $\alpha$  і  $\beta$  — вагові коефіцієнти для часових і асоціативних зв'язків ( $\alpha + \beta = 1$ ).

Ваги  $w_{ij}$  визначаються як середнє значення оцінок, отриманих від декількох мовних моделей, що входять до "рою віртуальних експертів":

$$w_{ij} = \frac{1}{n} \sum_{k=1}^n w_{ij}^k,$$

де  $n$  — кількість моделей,  $w_{ij}^k$  — оцінка моделі  $k$  для зв'язку між вузлами  $i$  та  $j$ .

Відповіді кожної моделі агрегуються через середнє значення оцінок. Для підвищення достовірності враховується частота підтвердження фактів різними моделями.

Відповіді різних мовних моделей порівнюються та узагальнюються. Особливу увагу приділяється випадкам розбіжностей у оцінках, які додатково перевіряються на відповідність фактам.

#### 4. Інтегральна оцінка достовірності

Інтегральна оцінка мережі проводиться з урахуванням вагових коефіцієнтів для часових і асоціативних зв'язків. Вагові коефіцієнти визначаються таким чином: для часових зв'язків  $\alpha = 0.7$ , для асоціативних  $\beta = 0.3$ . Для отримання інтегральної оцінки вводиться промпт:

*На основі наступних зв'язків і їхніх оцінок (часові та асоціативні) розрахуй інтегральну оцінку достовірності семантичної мережі:*

*Часові зв'язки: [перелік часових зв'язків у форматі "подія1;подія2;оцінка"].*

*Асоціативні зв'язки: [перелік асоціативних зв'язків у форматі "подія;поняття;оцінка"].*

*Врахуй вагові коефіцієнти: для часових зв'язків  $\alpha = 0.7$ , для асоціативних  $\beta = 0.3$ .*

*Результат виведи у вигляді: "Інтегральна оцінка = значення".*

**Очікуваний результат має приблизно такий вигляд:**

*"Інтегральна оцінка = 0.91"*

**Для додаткового пояснення достовірності застосовується промпт:**

*Поясни, які зв'язки (часові чи асоціативні) найбільше вплинули на загальну оцінку.*

*Виведи детальний аналіз окремих оцінок.*

#### 5. Агрегована оцінка

Для отримання агрегованої оцінки на основі оцінок зв'язків від кількох мовних моделей (GPT-4, Llama-3, o1) застосовується промпт:

*Для наступних оцінок зв'язків від кількох мовних моделей:*

*[перелік оцінок у форматі "подія1;подія2;оцінка моделі1, оцінка моделі2, оцінка моделі3"]*

*Розрахуй середню оцінку кожного зв'язку, а також загальну інтегральну оцінку мережі.*

Результат виведи у форматі:

- "Середня оцінка зв'язку: подія1;подія2;середнє значення".

- "Інтегральна оцінка мережі = значення".

**Очікуваний результат може мати такий вигляд:**

"Середня оцінка зв'язку: Народження;Навчання;0.95"

"Середня оцінка зв'язку: Навчання;Початок кар'єри;0.9"

"Інтегральна оцінка мережі = 0.92"

Приклад застосування

Для демонстрації методики було обрано біографію відомого вченого математика Леонарда Ейлера з Вікіпедії. Застосовано спрощену методику, що охоплює етапи формування семантичної мережі, і оцінки достовірності окремих зв'язків і вузлів і агрегації результатів різних LLM.

**Для формування семантичної мережі у цьому випадку застосовувались промпти для виділення основних подій біографії та побудова часових зв'язків між ними, а також знаходження понять та формування асоціативних зв'язків між ними та поняттями.**

Промпт для виділення основних подій біографії Леонарда Ейлера та побудова часових зв'язків між ними має вигляд:

*Візьми текст біографії Леонарда Ейлера. Визнач основні події у хронологічному порядку. Формулювати події коротко, до 5 слів. Виведи результат у форматі: "подія1;подія2", де "подія1" відбулася раніше за "подія2". Потім "подія2;подія3", потім "подія3;подія4". Тобто перша подія нового рядка відповідає другій події попереднього рядка". Приклади послідовності: "народився; школа"; "школа; інститут". І так далі. Кожна пара подій - окремим рядком. Ось текст біографії: ...*

У відповідь на цей промпт LLM видають приблизно таке:

*народився; навчання в гімназії*

*навчання в гімназії; навчання в університеті*

*навчання в університеті; отримання наукового звання магістра*

*отримання наукового звання магістра; відмова від теології*

*відмова від теології; професорська посада в Санкт-Петербурзі*

*професорська посада в Санкт-Петербурзі; знайомство з Християном Гольдбахом*

*знайомство з Християном Гольдбахом; обіймання посади професора фізики*

*обіймання посади професора фізики; обіймання посади професора математики*

*обіймання посади професора математики; осліплення на одне око*

*осліплення на одне око; запрошення до Берлінської академії*

*запрошення до Берлінської академії; повернення до Санкт-Петербурга*

*повернення до Санкт-Петербурга; остаточне осліплення*

*остаточне осліплення; смерть через крововилив у мозок*

Для додання асоціативних зв'язків між подіями та поняттями, що допомагають зрозуміти контекст кожної події формується промпт:

*На основі тексту біографії Леонарда Ейлера знайди концепти, пов'язані з кожною подією. Виведи результат у форматі: "подія;поняття". Події було визначено вище: народився; навчання в гімназії навчання в гімназії; навчання в університеті навчання в університеті; отримання наукового звання магістра*

Результат опрацювання такого промпту має такий вигляд:

*народився; Швейцарія, Базель, священик*

*навчання в гімназії; математика, Йоганнес Буркгардт*



навчання в університеті; університет Базеля, Йоганн Бернуллі  
отримання наукового звання магістра; філософські погляди, Ньютон, Декарт  
відмова від теології; теологія, наука  
професорська посада в Санкт-Петербурзі; університет Санкт-Петербурга, Даніель Бернуллі  
знайомство з Християном Гольдбахом; теорія чисел, Християн Гольдбах  
обіймання посади професора фізики; фізика, професор  
обіймання посади професора математики; математика, Даніель Бернуллі  
осліплення на одне око; втрата зору  
запрошення до Берлінської академії; Берлінська академія, Фрідріх Великий  
повернення до Санкт-Петербурга; Санкт-Петербург, наука  
остаточне осліплення; повна втрата зору  
смерть через крововилив у мозок; смерть, медичний стан

Таким чином із біографії визначено ключові події та побудовано мережу з вузлів і зв'язків. Візуалізацію цієї мережі можна побачити на Рис. 1.

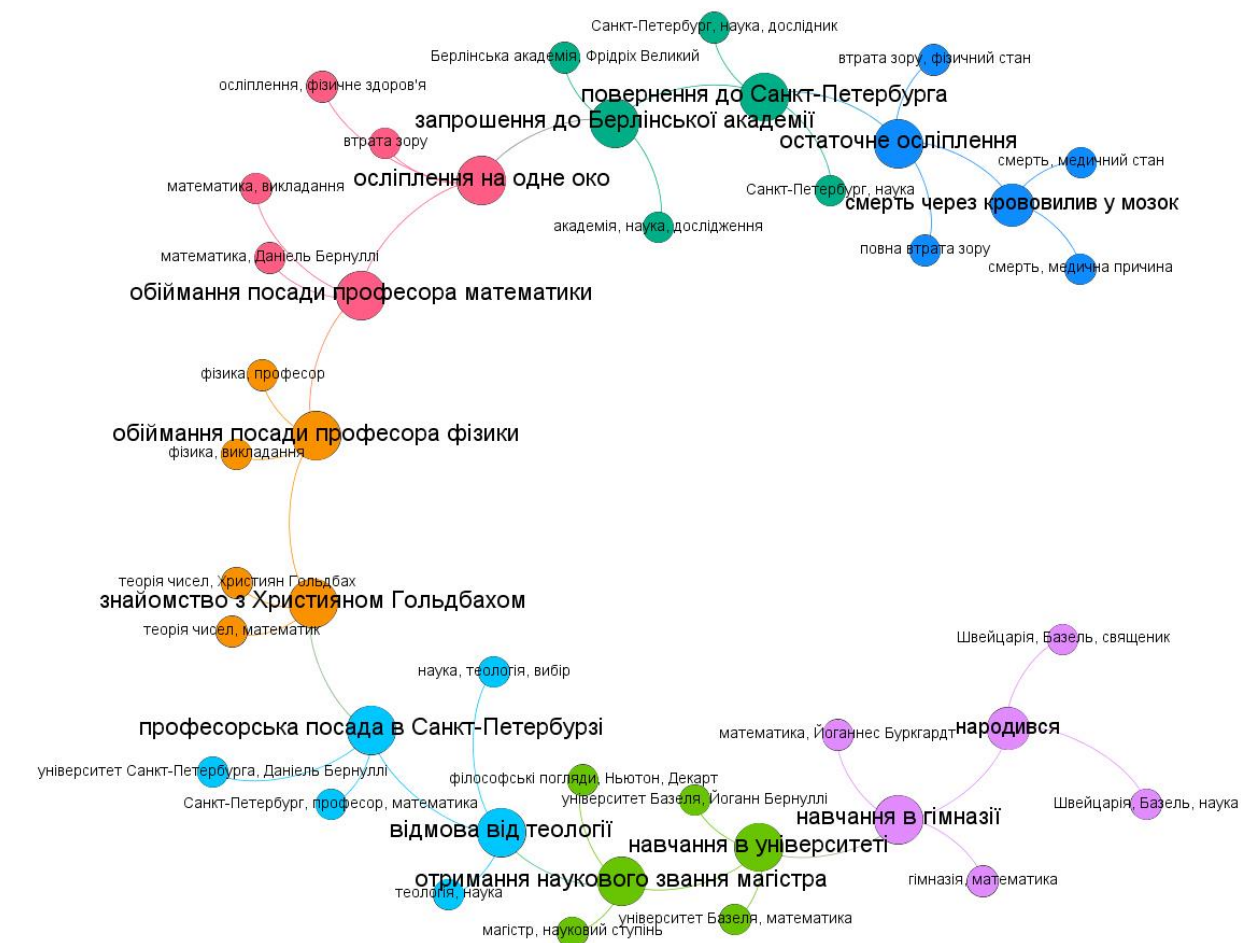


Рисунок 1 – Семантична мережа, що відповідає наведеній у Вікіпедії біографії Леонарда Ейлера

Оцінка достовірності для часових зв'язків проводиться на основі логіки хронології подій, а для асоціативних зв'язків — на основі коректності відображення поняття у контексті події. Для здійснення такої оцінки для кожної з LLM здійснюється звернення шляхом введення промпта:

Для наведеної мережі, де вузлами є події і концепти із біографії людини чисельно в діапазоні [0,1] оціни її загальну достовірність, узагальнюючи достовірність зв'язків між

подіями, концептами, а також концептами та подіями. Після цього надай коротке пояснення. Ось мережа:

*народився; навчання в гімназії  
навчання в гімназії; навчання в університеті  
навчання в університеті; отримання наукового звання магістра  
отримання наукового звання магістра; відмова від теології  
відмова від теології; професорська посада в Санкт-Петербурзі  
професорська посада в Санкт-Петербурзі; знайомство з Християном Гольдбахом  
знайомство з Християном Гольдбахом; обіймання посади професора фізики  
І т. і.*

У відповідь системи LLM надають свої оцінки достовірності, наприклад, DeepSeeс видає таке:

*Для оцінки загальної достовірності семантичної мережі, яка складається з подій та концептів, ми використовуємо зважене середнє значення достовірностей всіх зв'язків у мережі. Кожен зв'язок має свій ваговий коефіцієнт, який залежить від типу зв'язку (часовий-спрямований або асоціативний).*

*Вагові коефіцієнти:*

- Часові-спрямовані зв'язки:  $w_{time}=0.7$*
- Асоціативні зв'язки:  $w_{assoc}=0.3$*

*Оцінка достовірності зв'язків:*

**1. Часові-спрямовані зв'язки:**

- народився; навчання в гімназії:  $P1=0.9$*
- навчання в гімназії; навчання в університеті:  $P2=0.8$*
- навчання в університеті; отримання наукового звання магістра:  $P3=0.9$*
- отримання наукового звання магістра; відмова від теології:  $P4=0.7$*
- відмова від теології; професорська посада в Санкт-Петербурзі:  $P5=0.8$*
- професорська посада в Санкт-Петербурзі; знайомство з Християном Гольдбахом:  $P6=0.7$*
- знайомство з Християном Гольдбахом; обіймання посади професора фізики:  $P7=0.8$*
- обіймання посади професора фізики; обіймання посади професора математики:  $P8=0.9$*
- обіймання посади професора математики; осліплення на одне око:  $P9=0.6$*
- осліплення на одне око; запрошення до Берлінської академії:  $P10=0.7$*
- запрошення до Берлінської академії; повернення до Санкт-Петербурга:  $P11=0.8$*
- повернення до Санкт-Петербурга; остаточне осліплення:  $P12=0.7$*
- остаточне осліплення; смерть через крововилив у мозок:  $P13=0.9$*

**2. Асоціативні зв'язки:**

- народився; Швейцарія, Базель, священник:  $P14=0.8$*
- навчання в гімназії; математика, Йоганнес Буркгардт:  $P15=0.9$*
- навчання в університеті; університет Базеля, Йоганн Бернуллі:  $P16=0.8$*
- отримання наукового звання магістра; філософські погляди, Ньютон, Декарт:  $P17=0.7$*
- відмова від теології; теологія, наука:  $P18=0.8$*
- професорська посада в Санкт-Петербурзі; університет Санкт-Петербурга, Даніель Бернуллі:  $P19=0.7$*
- знайомство з Християном Гольдбахом; теорія чисел, Християн Гольдбах:  $P20=0.8$*
- обіймання посади професора фізики; фізика, професор:  $P21=0.9$*
- обіймання посади професора математики; математика, Даніель Бернуллі:  $P22=0.8$*

- осліплення на одне око; втрата зору: P23=0.7
- запрошення до Берлінської академії; Берлінська академія, Фрідріх Великий: P24=0.8
- повернення до Санкт-Петербурга; Санкт-Петербург, наука: P25=0.7
- остаточне осліплення; повна втрата зору: P26=0.9
- смерть через крововилив у мозок; смерть, медичний стан: P27=0.8
- народився; Швейцарія, Базель, наука: P28=0.9
- навчання в гімназії; гімназія, математика: P29=0.8
- навчання в університеті; університет Базеля, математика: P30=0.9
- отримання наукового звання магістра; магістр, науковий ступінь: P31=0.8
- відмова від теології; наука, теологія, вибір: P32=0.7
- професорська посада в Санкт-Петербурзі; Санкт-Петербург, професор, математика: P33=0.8
- знайомство з Християном Гольдбахом; теорія чисел, математик: P34=0.9
- обіймання посади професора фізики; фізика, викладання: P35=0.8
- обіймання посади професора математики; математика, викладання: P36=0.9
- осліплення на одне око; осліплення, фізичне здоров'я: P37=0.7
- запрошення до Берлінської академії; академія, наука, дослідження: P38=0.8
- повернення до Санкт-Петербурга; Санкт-Петербург, наука, дослідник: P39=0.7
- остаточне осліплення; втрата зору, фізичний стан: P40=0.9
- смерть через крововилив у мозок; смерть, медична причина: P41=0.8

Обчислення загальної достовірності:

$$P(G) = \left( \sum_{i=1}^{13} w_{time} \cdot P_i + \sum_{j=14}^{41} w_{assoc} \cdot P_j \right) / 41 \approx 0.76.$$

Загальна достовірність семантичної мережі становить приблизно 0.76, що вказує на досить високу достовірність біографії. Більшість зв'язків мають високу ймовірність достовірності, особливо часові-спрямовані зв'язки, які відображають хронологічну послідовність подій. Асоціативні зв'язки також мають високу достовірність, що свідчить про релевантність концептів до подій.

#### Висновки

У цій роботі була розроблена методологія та створена інформаційна технологія для виявлення та верифікації неправдивих фактів у біографіях Вікіпедії, що базується на використанні великих мовних моделей і семантичного нетворкінгу. Запропонований підхід дозволяє:

1. Побудувати детальну семантичну мережу біографії, орієнтовану на хронологічну послідовність подій та додавання релевантних концептів.
2. Оцінити достовірність окремих зв'язків та всієї мережі в цілому за допомогою промптів.
3. Використовувати різні великі мовні моделі для незалежної оцінки біографії та формування "рою віртуальних експертів".
4. Узагальнити результати оцінки різних моделей та надати загальну оцінку достовірності біографії.

Наукова новизна роботи полягає у наступному:

1. Вперше запропоновано методику побудови семантичної мережі біографій з двома типами зв'язків між подіями та концептами.
2. Уперше застосовано концепцію "рою віртуальних експертів" для верифікації текстових даних у біографічних статтях.

3. Визначено цифрові критерії оцінки достовірності як окремих зв'язків, так і всієї мережі в цілому
4. Розроблено методологію оцінки достовірності, що поєднує LLM та семантичний нетворкінг.
5. Впроваджено модель інтегральної оцінки, яка враховує вагові коефіцієнти часових та асоціативних зв'язків.

Крім того, запропонований підхід може бути використаний для автоматизованої верифікації фактів у відкритих джерелах. Методика дозволяє швидко ідентифікувати потенційно маніпулятивні дані, що є актуальним для журналістики, історичних досліджень і боротьби з дезінформацією.

Унікальність методу також полягає в застосуванні агрегованої оцінки достовірності, що враховує результати різних мовних моделей, що забезпечує більш високу точність та стійкість до можливих помилок окремих моделей. Цей підхід дозволяє уникнути упередженості, що може виникнути через специфічні обмеження кожної окремої моделі, та забезпечує більш надійну перевірку фактів.

Запропонована методика демонструє дієвість у виявленні потенційних неправдивих фактів, підвищуючи достовірність інформації у відкритих джерелах. Вона відкриває перспективи для автоматизації перевірки інформації у відкритих джерелах, зокрема Вікіпедії, та може бути адаптована для аналізу інших видів текстових даних.

#### **Список посилань.**

1. Joshi N., Spezzano F., Green M., & Hill E. (2020, April). *Detecting undisclosed paid editing in wikipedia*. In *Proceedings of The Web Conference 2020* (pp. 2899-2905). DOI: 10.1145/3366423.3380055
2. Gao Fan, Hang Jiang, Rui Yang, Qingcheng Zeng, Jinghui Lu, Moritz Blum, Tianwei She, Yuang Jiang, and Irene Li. "Evaluating large language models on wikipedia-style survey generation." In *Findings of the Association for Computational Linguistics ACL 2024*, pp. 5405-5418. 2024. DOI: 10.18653/v1/2024.findings-acl.321
3. Ding N., Mayer W., Geng Y., Duan Y., & Feng Z. (2023, December). *Generative Semantic Modeling for Structured Data Source with Large Language Model*. In *2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*. pp. 1148-1152. IEEE. DOI: 10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00164
4. Lande D., LStrashnoy L. *GPT Semantic Networking: A Dream of the Semantic Web - The Time is Now*. - Kyiv: Engineering, 2023. - 168 p. ISBN 978-966-2344-94-3
5. Wu J., Guo J., & Hooi B. (2024, August). *Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks*. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3367-3378). DOI: 10.1145/3637528.3671977
6. Xie Bingbing, Xiaoxiao Ma, Xue Shan, Amin Beheshti, Jian Yang, Hao Fan, and Jia Wu. "Multiknowledge and LLM-Inspired Heterogeneous Graph Neural Network for Fake News Detection." *IEEE Transactions on Computational Social Systems* (2024). DOI: 10.1109/TCSS.2024.3488191
7. Lande D., Humeniuk O. *Content Reconstruction: The Evolution of Texts through Semantic Networks and LLMs/ Available at SSRN Preprint*, Sep. 09, 2024). - 11 p. DOI: <http://dx.doi.org/10.2139/ssrn.4951516>
8. Lande D., Strashnoy L. *Swarm of Virtual Experts in the implementation of Semantic Networking*. *ResearchGate Preprint*, October 2024. DOI: 10.13140/RG.2.2.16686.11845
9. Ланде Д., Страшной Л., Свобода І. Застосування концепції рою віртуальних експертів при формуванні семантичних мереж. Міжнародна наукова-технічна конференція "Інтелектуальні технології лінгвістичного аналізу": Тези доповідей. - Київ: НАУ, 2024. - С. 24. URL: <http://dwl.kiev.ua/art/nau2024c/>
10. Feher A., Lande D. *Defined AI semantic networking in cybersecurity // Intelligent Solutions-S: Proceedings of the International Symposium, September 28, 2023, Kyiv-Uzhorod, Ukraine*. Kyiv: Publishing House "Caravela", 2023. pp. 21-22. ISBN 978-966-801-916-6

**Мазурець О.В., Овчарук О.М.,** (Хмельницький національний університет, м. Хмельницький, Україна)

## ДІАГНОСТУВАННЯ ПРОЯВІВ ПОСТТРАВМАТИЧНОГО СТРЕСОВОГО РОЗЛАДУ ЗА НЕЙРОМЕРЕЖЕВИМ АНАЛІЗОМ ТЕКСТОВОГО КОНТЕНТУ

**Анотація:** Запропоновано підхід до діагностування проявів посттравматичного стресового розладу за нейромережєвим аналізом текстового контенту, який забезпечує трансформацію вхідних даних у вигляді тексту в результат у вигляді оцінки ймовірності проявів ПТСР у контенті користувачів. Розроблена архітектура нейронної мережі показала значне покращення результатів порівняно з існуючими аналогами. Після навчання моделі на запропонованому наборі даних було отримано точність понад 85%.

**Ключові слова:** посттравматичний стресовий розлад, нейронна мережа, обробка природної мови

**Abstract:** Approach to diagnosing manifestations of post-traumatic stress disorder using neural network analysis of textual content is proposed, which provides the transformation of input data in the form of textual data into a result in the form of an assessment of the probability of manifestations of PTSD in user content. The developed neural network architecture showed a significant improvement in results compared to existing analogues. After training the model on the proposed data set, an accuracy of over 85% was obtained.

**Keywords:** post-traumatic stress disorder, neural network, natural language processing

### Постановка проблеми

На тлі останніх подій, які викликають значний стрес, таких як війни, катастрофи чи соціальні конфлікти, спостерігається зростання випадків посттравматичного стресового розладу (ПТСР) серед населення [1, 2]. Традиційні підходи до діагностики часто виявляються недостатньо ефективними, оскільки потребують безпосереднього контакту з пацієнтом та тривалого часу для оцінки його стану.

У зв'язку з тим, що текстова інформація стає дедалі доступнішою, виникає необхідність у створенні методології формування навчальних наборів даних, яка забезпечить високу точність та чутливість моделей до текстових проявів ПТСР [3]. Це дозволить не лише підвищити ефективність діагностики, але й сприятиме своєчасному втручання та підтримці осіб, які пережили травматичні події. В умовах швидких змін у соціально-психологічному середовищі використання нейронних мереж для виявлення ознак ПТСР стає важливим і актуальним завданням, яке спрямоване на забезпечення цифрового психологічного добробуту суспільства.

Таким чином, актуальним є питання автоматизації діагностики ПТСР на основі текстового контенту, забезпечуючи при цьому стійкість до помилок у розрізненні ПТСР та інших психічних розладів. Цього можна досягти за допомогою нейронних мереж шляхом створення навчального набору даних, що включає тексти з яскраво вираженими проявами ПТСР як цільову категорію, а також тексти з іншими психічними розладами та без них – як контрольну групу.

### Аналіз останніх публікацій

Наразі широко застосовуються такі цифрові інструменти, як віртуальна реальність, нейрофідбек, носимі пристрої, мобільні додатки та машинне навчання для діагностики та лікування симптомів ПТСР [4]. Аналіз досліджень показав, що віртуальна реальність є ефективним методом лікування бойового ПТСР, тоді як результати використання мобільних додатків і нейрофідбеку залишаються обмеженими через невеликі вибірки та методологічну неоднорідність. Моделі машинного навчання демонструють потенціал у виявленні симптомів ПТСР за даними зі смартфонів і носимих пристроїв, однак для їх клінічного застосування потрібна стандартизація процедур збору даних і моделювання.

Дослідження [5] розглядає мову як можливий діагностичний біомаркер ПТСР, використовуючи дані від 148 осіб, які стали жертвами терактів у Парижі 13 листопада 2015

року. Інтерв'ю, проведені через 5-11 місяців після подій, охоплювали респондентів зі схожими соціально-економічними умовами. У дослідженні застосовувалася міждисциплінарна методологія, яка поєднувала психіатрію, лінгвістику та обробку природної мови для встановлення зв'язку між мовними особливостями та ПТСР.

Результати показали, що клінічний психіатр досяг AUC 0,72 при діагностиці ПТСР, що зіставно з результатами анкетування (AUC  $\approx$  0,80). Модель машинного навчання показала AUC 0,69, а модель глибокого навчання – AUC 0,64. У дослідженні також враховувалися змішувальні фактори, що впливають на результати, встановлювалися зв'язки між мовними особливостями та симптомами DSM-5, а також поєднувалися автоматизовані методи з якісним аналізом.

### Мета роботи та постановка завдань

Метою роботи є діагностування проявів посттравматичного стресового розладу за нейромережевим аналізом текстового контенту, що забезпечує перетворення вхідних даних у вигляді тексту в результат у вигляді оцінки ймовірності проявів ПТСР у контенті користувачів.

### Виклад основного матеріалу

Метод діагностування проявів посттравматичного стресового розладу за нейромережевим аналізом текстового контенту базується на навчанні нейронної мережі з використанням спеціально створеного навчального набору даних. Цей набір забезпечує високу стійкість до помилкового виявлення ПТСР через змішування з іншими психічними розладами. Для аналізу текстів була застосована нейронна мережа, що включає квантовий шар. Схема архітектури мережі для діагностування проявів посттравматичного стресового розладу за аналізом текстового контенту представлена на рис. 1.

Запропонований підхід до діагностування проявів посттравматичного стресового розладу за нейромережевим аналізом текстового контенту забезпечує трансформацію вхідних даних у вигляді тексту в результат у вигляді оцінки ймовірності проявів ПТСР у контенті користувачів. На першому етапі здійснюється попередня обробка текстів, включаючи їх токенизацію. Тексти перевіряються на довжину і наявність даних. При цьому зберігаються розділові знаки, емодзі та інші елементи, які можуть містити значущу інформацію для виявлення ПТСР. Процес токенизації виконується із застосуванням токенизатора, що використовувався під час навчання моделі.

На другому етапі проводиться аналіз тексту за допомогою контекстно-орієнтованої нейромережі на основі архітектури трансформера, спрямованої на виявлення проявів ПТСР.

Вхідний шар (Input) моделі визначає формат даних і має розмірність 2000, що відповідає довжині вектора слів для аналізу. На рисунку цей шар не відображено. Подальші шари включають щільні (Dense) шари, що з'єднують усі нейрони попереднього шару з кожним нейроном поточного. Dropout-шар випадково деактивує певну кількість нейронів у процесі навчання, що зменшує ризик перенавчання моделі, зокрема для даного дослідження використовується значення 70%.

Вихідний шар Dense із одним нейроном використовується для задачі класифікації на два класи. Застосовується функція активації Sigmoid, що стискає результат у межах від 0 до 1, що ідеально підходить для моделювання ймовірності проявів ПТСР.

Для забезпечення точності діагностики та зменшення ймовірності змішування ПТСР із симптомами інших психічних захворювань було створено навчальний набір даних. У ньому цільова категорія містить тексти з високим рівнем проявів ПТСР, а контрольна – включає як тексти без ознак ПТСР, так і тексти, пов'язані з іншими психічними розладами.

Формування цього набору здійснювалося шляхом комбінування та відбору даних із наявних баз [6], таких як «AyaPTSD» [7] та «HumanStressPrediction» [8], розміщених на платформі Kaggle.

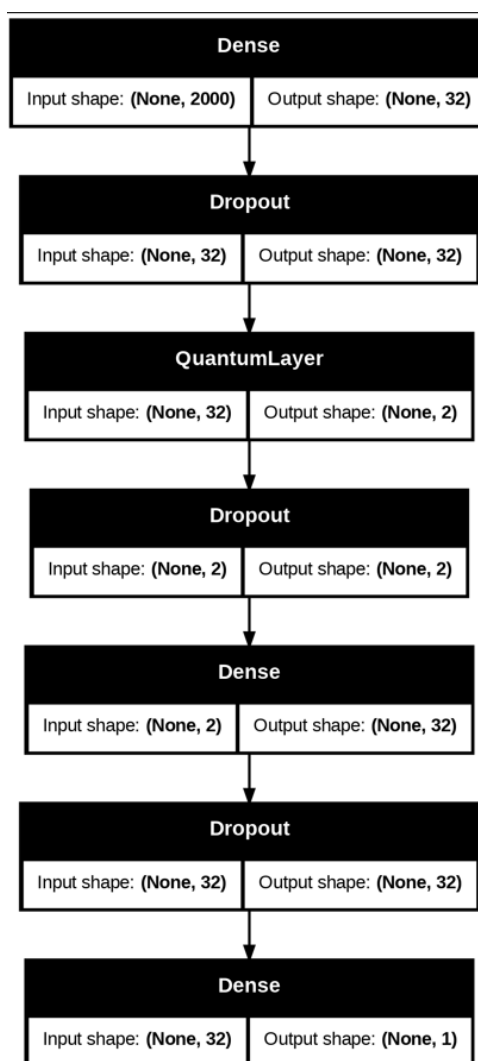


Рис. 1. Архітектура нейромережі для аналізу текстового контенту для діагностування проявів ПТСР

Ефективність методу діагностування проявів посттравматичного стресового розладу за нейромережним аналізом текстового контенту була оцінена за допомогою хмарної платформи Google Colab. На рисунках 2 і 3 продемонстровані ROC-криві для двох варіантів параметрів Dropout – 60% і 75%.

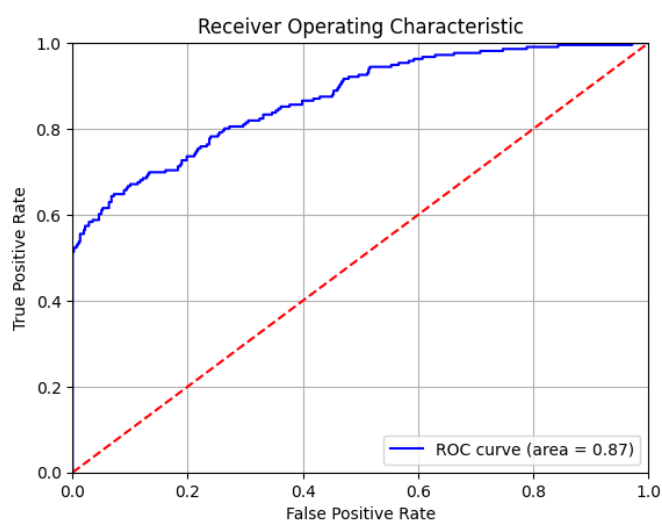


Рис. 2. ROC-криві нейронної мережі, відсоток інгібуючих нейронів 60

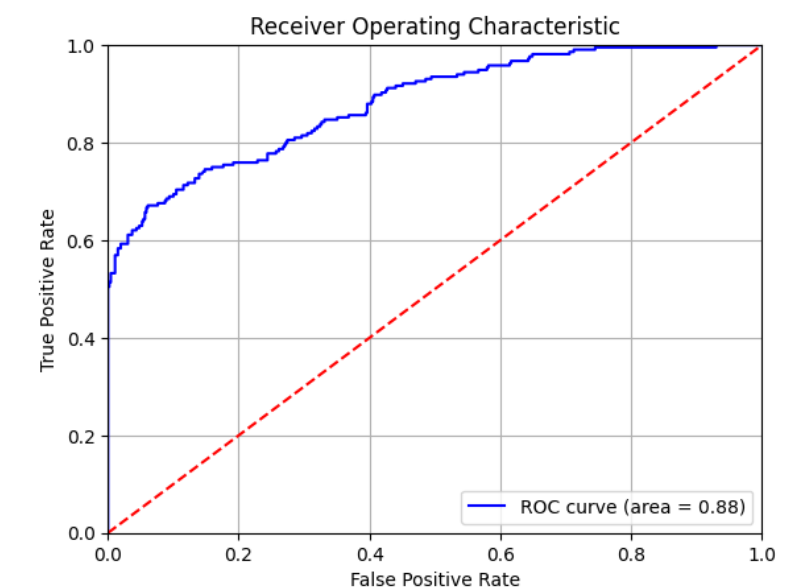


Рис. 3. ROC-криві нейронної мережі, відсоток інгібуючих нейронів 75

Після навчання моделі на запропонованому наборі даних були отримані такі результати: точність (Accuracy) – 0.85, точність передбачень (Precision) – 0.87, і площа під кривою (AUC) – 0.88. Навчання здійснювалося протягом 100 епох із застосуванням механізму ранніх зупинок для уникнення перенавчання та збереження найкращої версії моделі. Розмір навчальної партії (batch size) становив 64, а словник охоплював 2000 токенів.

## ВИСНОВКИ

Розроблена архітектура нейронної мережі показала значне покращення результатів порівняно з існуючими аналогами. Це особливо важливо для задач виявлення ПТСР у текстовому контенті. Встановлено підвищення точності і зменшення ризику сплутування з іншими психічними розладами. Однак метод діагностування проявів посттравматичного стресового розладу за нейромережовим аналізом текстового контенту має певні обмеження: він застосовується лише до текстів англійською мовою. Це обмеження можна подолати шляхом розширення навчального набору даних текстами іншими мовами.

Подальші дослідження будуть зосереджені на розширенні набору навчальних даних і тестуванні альтернативних нейромережових архітектур для підвищення ефективності моделі та точності виявлення симптомів ПТСР.

## Список посилань.

11. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. *Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network*. *CEUR Workshop Proceedings*, 2024, vol. 3688, pp. 16-28. <https://doi.org/10.31110/COLINS/2024-3/002>
12. Park, A. H., Patel, H., Mirabelli, J., Eder, S. J., Steyrl, D., Lueger-Schuster, B., Scharnowski, F., O'Connor, C., Martin, P., Lanius, R. A., McKinnon, M. C., & Nicholson, A. A. (2023). *Machine learning models predict PTSD severity and functional impairment: A personalized medicine approach for uncovering complex associations among heterogeneous symptom profiles*. *Psychological Trauma: Theory, Research, Practice, and Policy*. Advance online publication. <https://doi.org/10.1037/tra0001602>
13. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. *Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network*. *CEUR Workshop Proceedings*, 2023, vol. 3387, pp. 344–356.
14. Browning, L.; Rashid, I.; Javanbakht, A. *The Current State of Digital Technologies for the Treatment and Management of PTSD- A Look into the Future of Psychiatry*. *Preprints* 2024, 2024091313. <https://doi.org/10.20944/preprints202409.1313.v1>



15. Quillivic, R., Gayraud, F., Auxéméry, Y. et al. *Interdisciplinary approach to identify language markers for post-traumatic stress disorder using machine learning and deep learning*. *Sci Rep* 14, 12468 (2024). <https://doi.org/10.1038/s41598-024-61557-7>
16. Мазурець О.В., Овчарук О.М. Підхід до збільшення розмірності вхідних даних для нейромережевого прогнозування значень показників епідеміологічної небезпеки. *Матеріали VIII Міжнародної науково-практичної конференції «Перспективи сучасної науки: теорія і практика»*. 16-18.09.2024. Львів – 2024. с. 192-198.
17. Набір даних «AyaPTSD». URL: <https://www.kaggle.com/datasets/abdelrahmanahmed3/aya-ptsd>
18. Набір даних «HumanStressPrediction». URL: <https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction/data>

**Мартинюв А.Ю., Радюк П.М.** (Хмельницький національний університет, м. Хмельницький, Україна)

## ШТУЧНИЙ ІНТЕЛЕКТ ЯК ОСНОВА ДЛЯ РОЗПІЗНАВАННЯ БПЛА ПО РАДІОСИГНАТУРИ

**Анотація:** У статті представлені пару простих нейронних мереж для розпізнавання радіосигналів по типу модуляції чи маніпуляції. Вибравши кращу із них ми покращили її підналаштуванням кількості рівнів шарів та їх параметри. Також будемо звертати увагу на затребуваний час для навчання та результати досліджень.

**Ключові слова:** маніпуляції, покращення, підналаштування, кількість шарів, параметри, час навчання, результати досліджень.

**Abstract:** The article presents a pair of simple neural networks for recognizing radio signals by the type of modulation or manipulation. Having chosen the best of them, we will improve it by adjusting the number of layer levels and their parameters. We will also pay attention to the required time for training and research results.

**Keywords:** manipulations, improvements, fine-tuning, number of layers, parameters, training time, research results.

### Вступ.

Через велику кількість виробників та типів ворожих БПЛА постає проблема розпізнати свій-чужий БПЛА на фоні активних бойових дій. Це відбувається за умов, коли переважна більшість сучасної розвідки та ураження відбувається за рахунок FPV-дронів [1–3]. Визначити тип БПЛА можна і за спектральним радіосигналом оператора [5] чи відеосигналу самого БПЛА [6–9].

### Запропонований підхід.

Запропонований метод виявлення ворожого БПЛА запропонований вперше. Він охоплює в собі цифровий радіоприймач, який приймає сигнали керування оператора чи відеосигнал самого БПЛА, комп'ютер чи міні комп'ютер на якому буде розташована сама нейронна мережа із розпізнаванням типу радіосигналу та індикатор, який буде свідчити про наявність БПЛА.

У процесі пошуку нейронних мереж для розпізнавання радіосигналів було добуто основних п'ять типів нейронних мережможливих для розпізнавання радіосигналів. Із них будемовибрати оптимальнішу і вже покращувати її зміною кількості шарів та їх параметрів методом перебору.

Розпізнати БПЛА планується за наступними кроками:

**Блок 1.** Навчання за набором даних.

*Крок 1.1* Запускаємо нейронні мережі та тренуємо згідно нашого набору даних.

*Крок 1.2* виводимо результат тренувань у вигляді «Матриці плутанини». Якщо результати задовільні – переходимо до блоку 2

**Блок 2.** Отримуємо сигнал для порівняння

*Крок 2.1* Отримуємо радіосигнал з ефіру керування оператора чи відеосигнал із самого БПЛА

*Крок 2.2* Приводимо сигнал у відповідну форму для нейромережі

**Блок 3.** Виведення результатів для прийняття рішень

*Крок 3.1.* Порівняння результатів із базою даних по типам БПЛА (по характеристиках модулів, які можуть використовуватися в них)

*Крок 3.2.* Виведення інформації із ймовірнісним застосування певного типу БПЛА)

## Опис даних.

За основу було взято dataset «RadioML 2016.10A» 11 модуляцій, було добавлено ще один тип модуляції «протокол LoRa» на якому і працюють важкі для придушення БПЛА.

Утворений новий dataset має наступні значення:

Кількість зразків модуляцій 12;

Кількість зразків в одному виді 1000;

Розмірність компонента 128x128;

Кількість рівнів SNR 20;

Загальна кількість зразків 240000.

## Процедура навчання моделі.

Навчаємо неймережу за набором даних. Після тренувань прогнали ще раз набір даних, який буде на всі 100% складатися із Навчання і тільки наш вхідний сигнал записаний цифровим приймачем подаватимемо на вхід на розпізнавання до якої модуляції відноситься цей записаний радіосигнал. На виході отримаємо ймовірнісні значення цього сигналу у відсотковому відношенні до типу модуляцій. Пізніше порівнюємо характеристики існуючих модулів передачі радіосигналів з нашим розпізнаним сигналом і отримуємо ймовірнісне застосування певного типу передавача на досліджуваному БПЛА.

Описані 5 структур нейронних мереж які будемо досліджувати:

1. BASIC = Conv2D(64, size=(1, 3), activation='relu', input\_shape=(H, W, C)), Dropout(0.5), Conv2D(16, size=(2, 3), activation='relu', Dropout(0.5), Flatten(), Dense(128, activation='relu'), Dense(len(modulation), activation='softmax'), Dropout(0.5)]

2. DEEP = [Conv2D(64, kernel\_size=(1, 3), activation='relu', input\_shape=(H, W, C)), Dropout(0.6), Conv2D(64, kernel\_size=(2, 3), activation='relu', Dropout(0.6), Conv2D(80, kernel\_size=(1, 3), activation='relu', Dropout(0.6), Conv2D(80, kernel\_size=(1, 3), activation='relu', Dropout(0.6), Flatten(), Dense(128, activation='relu'), Dropout(0.5), Dense(len(modulation), activation='softmax')]

3. LSTM = [ZeroPadding2D(0, 2), Conv2D(64, (1, 5), activation='relu', Dropout(0.2), ZeroPadding2D((0, 2), data\_format='channels\_last'), Conv2D(64, (1, 5), activation='relu', Dropout(0.2)(conv\_2), Concatenate(axis=2)([drop\_1, drop\_2]), list(np.shape(merge)), concat\_n, concat\_h, concat\_w, units = np.shape(merge), dimensions = int(concat\_h)\*int(concat\_w), Reshape((units, dimensions)), Bidirectional(LSTM(64)), Dense(128, activation='relu'), Dense(len(modulation), activation='softmax')]

4. CNN = [ZeroPadding2D(0, 2), Conv2D(64, (2, 3), activation='relu', data\_format='channels\_last', input\_shape=(H, W, C)), Dropout(0.5), Conv2D(80, (1, 3), activation='relu', data\_format='channels\_last'), Dropout(0.5), Flatten(), Dense(128, activation='relu'), Dense(len(modulation), activation='softmax')]

5. VTCNN2 = [Reshape([1]+in\_shp, input\_shape=in\_shp)), ZeroPadding2D((0, 2)), Convolution2D(64, (2, 3), activation="relu", initializer='glorot\_uniform'), Dropout(0.0005), ZeroPadding2D((0, 2)), odel.add(Dropout(0.0005), Flatten(), Dense(128, activation='relu', kernel\_initializer='he\_normal'), Dropout(0.0005), Dense(len(modulation), kernel\_initializer='he\_normal'), Activation('softmax') Reshape([len(modulation)])]

Нейронні мережі тренуємо у сукупності усіх рівнів SNR а не окремо. SNR = [-20, -18, -16, -14, -12, -10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10, 12, 14, 16, 18].

## Результати.

Для нейронної мережі типу «DEEP» на 50 епох приведені в рисунках 1 – по точності, рисунок 2 - по втрагах, 3 – матриця плутанини.

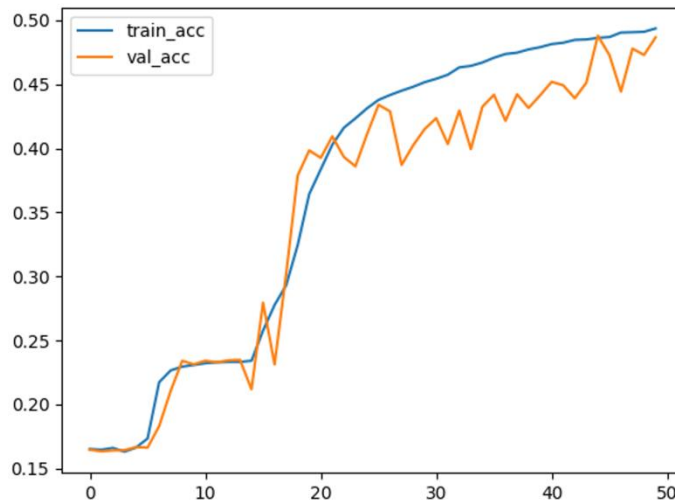


Рисунок 1 – Значення точності нейронної мережі «DEEP» протягом 50 - епох.

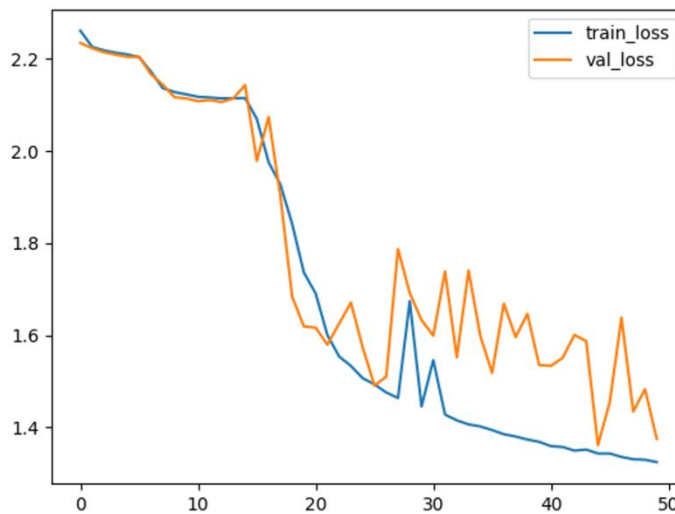


Рисунок 2 – Значення втрат нейронної мережі «DEEP» протягом 50 - епох.

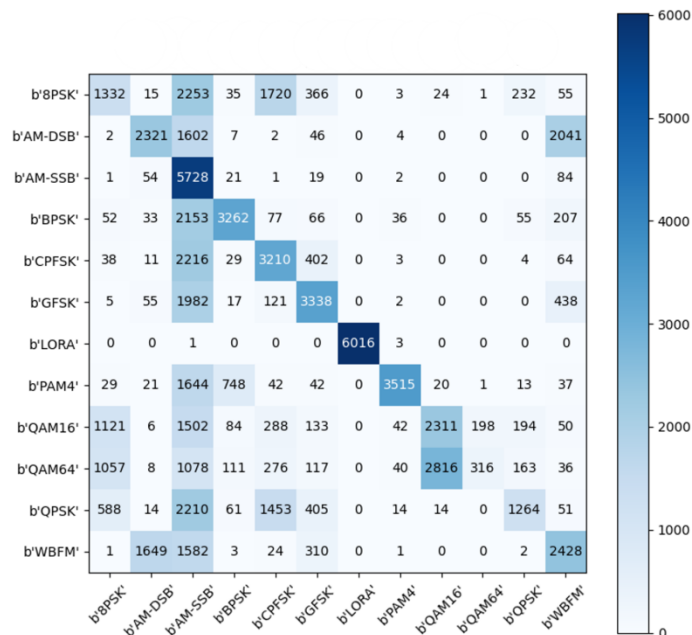


Рисунок 3 – Матриця плутанини нейронної мережі «DEEP» за 50 - епох.

Таблиця 1 – Типи нейронних мереж виражені у відсотковому представленні розпізнавань радіосигналів після тренувань за кількістю епох.

Тип мережі	BASIC	DEEP	LSTM	SNN	VTCNN2
5	8,05	16,54	16,49	16,11	34,06
10	8,13	-	31,97	16,19	48,06
20	8,06	-	54,61	16,15	47,55
50	9,3	44,49	56,5	20,13	52,10
500	-	-	-	-	50,46

Таблиця 2 – Типи нейронних мереж виражені у потребі в часі (секунди) для тренувань за кількістю епох.

Тип мережі	BASIC	DEEP	LSTM	SNN	VTCNN2
5	387	1206	1001	200	120
10	694	-	2085	346	278
20	1375	-	8109	762	514
50	7344	19698	18249	1844	5284
500	-	-	-	-	20303

Таблиця 3 – Показник затребуваного часу до її кількості епох, для тренувань по типу нейронної мережі.

Тип мережі	BASIC	DEEP	LSTM	SNN	VTCNN2
5	1,29	4,02	3,67	0,67	0,4
10	1,16	-	3,48	0,58	0,46
20	1,15	-	6,76	0,64	0,43
50	2,45	6,57	6,08	0,61	1,76
500	-	-	-	-	0,68

По результатах наведених в таблиці 1 бачимо, що «BASIC» тип нейронної мережі по точності розпізнавання радіосигналів досить мала а тому не підходить для подальших дослідів.

По результатах приведених в таблиці 2 бачимо, що «DEEP» тип нейронної мережі забирає багато часу для тренувань, а тому її відкидаємо із-за недостатньої потужності нашого обладнання.

Для нейронної мережі типу «LSTM» ми спостерігаємо геометричний ріст затребуваного часу від кількості епох (згідно таблиці 2) але і ріст розпізнаваності радіосигналу (згідно таблиці 1).

Для нейронної мережі типу «SNN», згідно таблиці 2, ми спостерігаємо лінійну рівність затребуваного часу для тренувань але відсоток досить малий. Можливо при тренуванні біля 1000 чи 5000 ми доб'ємося відсоткового значення більше за 75 але поки недостатня потужність нашого обладнання не дає змогу це перевірити.

Для нейронної мережі типу «VTCNN2», згідно таблиці 3, ми бачимо аномалію що це значення більше одного. Потрібно дослідити це детальніше.

Взявши за основу цей тип нейромережі такий як «VTCNN2» із налаштуванням шарів можна досягти кращих результатів. Ось пару змін шарів та результати обчислень на 50 епох представлені в таблиці 4.

Таблиця 4 – структури шарів для нейронної мережі.

Назва моделі	00	03	08	12	14	64	128	256	512
ZeroPadding2D((0, 2))	+	+	+	+	+	+	+	+	+
Convolution2D(64,(2,3),Activation="relu",initializer='glorot_uniform')	+	+	sigmoid	+	128	64	128	256	512
Dropout(0.0005)	+	+	+	+	+	+	+	+	+

ZeroPadding2D((0, 2))	+	+	+	+	+	+	+	+	+
Dropout(0.0005)	+		+		+	+	+	+	+
Flatten()	+		+		+	+	+	+	+
Dense (128, activation='relu', initializer='he_normal')	+		+		256	+	256	512	1024
Dropout (0.0005)		+	+	+	+	+	+	+	+
Flatten()		+	+	+	+	+	+	+	+
Dense(128, ctivation='sigmoid', initializer='he_normal')		+	+	+	256	+	256	512	1024
Dropout (0.0005)	+	+	+		+	+	+	+	+
Flatten()	+	+	+		+	+	+	+	+
Dense(128, activation='relu', initializer='he_normal')	+	+	+		256	+	256	512	1024
Dropout (0.0005)					+	+	+	+	+
Flatten()					+	+	+	+	+
Dense (128, activation='relu', initializer='he_normal')					256	+	256	512	1024
Dropout (0.0005)	+	+							
Flatten()	+	+							
Dense (128, activation='sigmoid', initializer='he_normal')	+	+							
Dropout (0.0005)	+	+	+	+	+	+	+	+	+
Dense (len(modulation), initializer='he_normal')	+	+	+	+	+	+	+	+	+
Activation ('softmax')	+	+	+	+	+	Soft plus	Soft plus	Soft plus	Soft plus
Reshape ([len(modulation)])	+	+	+	+	+	+	+	+	+
model.compile (loss='categorical_crossentropy', optimizer='adam')	+	+	+	+	+	+	+	+	+
model.build (input_shape=(None, H, W, C))	+	+	+	+	+	+	+	+	+
summary()	+	+	+	+	+	+	+	+	+
Результат розпізнавання у %	1887 52,37	6444 51,46	1334 8,38	7556 51,42	6872 53,88	133 40,78	306 44,44	4808 35,47	2670 33,01
Час затрачений для Навчання в секундах	1887	6444	1334	7556	6872	133	306	4808	2670

Визначили оптимальну для наших подальших напрацювань.

### Висновки.

За основу для нашої нейронної мережі візьмемо нейромережу із такою структурою шарів під назвою «00». Відсоток розпізнавання більший від 50 та затребуваний час середній відносно інших.

Подальша робота буде зосереджена на зборі характеристик передавачів для БПЛА та їхні особливості керування (частоти та типи протоколів, модуляції, можливість стрибків по частоті, можливість підстроювання сигналу до рівня шумів тощо).

### Список посилань.

1. A Survey of Traffic Prediction Based on Deep Neural Network: Data, Methods and Challenges / P. Cao et al. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Cham, 2022. P. 17–29. URL: [https://doi.org/10.1007/978-3-030-99191-3\\_2](https://doi.org/10.1007/978-3-030-99191-3_2)
2. Azzouz E. E., Nandi A. K. *Modulation Recognition Using Artificial Neural Networks. Automatic Modulation Recognition of Communication Signals*. Boston, MA, 1996. P. 132–176. URL: [https://doi.org/10.1007/978-1-4757-2469-1\\_5](https://doi.org/10.1007/978-1-4757-2469-1_5)
3. Cao G., Wu F., Zhao J. A Survey of Deep Neural Network Compression. *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*. Cham, 2021. P. 1442–1463. URL: [https://doi.org/10.1007/978-3-030-70665-4\\_157](https://doi.org/10.1007/978-3-030-70665-4_157)
4. Radiuk P.M. Application of a genetic algorithm to search for the optimal convolutional neural network architecture with weight distribution. *Herald of Khmelnytskyi National University. Technical sciences*. 2020. Vol. 281, no. 1. P. 7–11. URL: <https://doi.org/10.31891/2307-5732-2020-281-1-7-11>

5. Liu X., Yang D., Gamal A. E. *Deep neural network architectures for modulation classification*. 2017 51st Asilomar Conference on Signals, Systems, and Computers, m. Pacific Grove, CA, USA, 29 Oct. – 1 Nov. 2017 p. 2017. URL: <https://doi.org/10.1109/acssc.2017.8335483>
6. *Model Compression and Acceleration for Deep Neural Networks: The Principles, Progress, and Challenges* / Y. Cheng et al. *IEEE Signal Processing Magazine*. 2018. Vol. 35, no 1. C. 126–136. URL: <https://doi.org/10.1109/msp.2017.2765695>
7. *Human-in-the-loop approach based on MRI and ECG for healthcare diagnosis* / P. Radiuk et al. *Proceedings of the 5th International Conference on Informatics & Data-Driven Medicine : CEUR-Workshop Proceedings, Lyon, France, 18–20 November 2022* / ed. by N. Shakhovska et al. Aachen, 2022. P. 9–20. URL: <https://ceur-ws.org/Vol-3302/paper1.pdf>
8. *SNAP: An Efficient Sparse Neural Acceleration Processor for Unstructured Sparse Deep[-1pt] Neural Network Inference* / J.-F. Zhang et al. *IEEE Journal of Solid-State Circuits*. 2020. P. 1. URL: <https://doi.org/10.1109/jssc.2020.3043870>
9. *Survey of automatic modulation classification techniques: classical approaches and new trends* / O. A. Dobre et al. *IET Communications*. 2007. No. 1, no 2. P. 137. URL: <https://doi.org/10.1049/iet-com:20050176>

**Мельников О. Ю., Денисенко В. О.** (Донбаська державна машинобудівна академія, м. Краматорськ, Україна)

## ПРОГНОЗУВАННЯ ЗМІНИ РІВНЯ ЛІСОВОГО ВКРИТТЯ ОКРЕМОГО ЛІСНИЦТВА ЗА ДОПОМОГОЮ СУПУТНИКОВОГО СЕРВІСУ LANDSAT ТА ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

**Анотація:** Розглянуто проблему розрахунку рівня лісистості, у тому числі – прогнозування зміни лісового вкриття в окремому лісництві. Зроблено порівняння лісистості за 40 років (1984 – 2024) селища Співаківка Ізюмського району Харківської області з використанням супутникового сервісу Landsat. Для застосування методу штучних нейронних мереж сформовано перелік входних факторів, що містять показники на обраній ділянці у два попередні роки та ці ж показники на сусідніх ділянках (кожне із завантажених зображень обраного лісництва розподілено на 9 квадратів). Створено скрипт мовою програмування та аналізу даних R, який виконує розрахунки, дозволяє визначити найкращу архітектуру нейронної мережі та більш ефективний метод її навчання для певного набору даних. Наведено розрахунок динаміки вирубки на обраному квадраті (прогноз на останній рік забезпечує похибку в 9%). Після численних запусків скрипту з'ясовано, що найкращий результат забезпечує перцептрон з одним прихованим шаром з двома нейронами у ньому.

**Ключові слова:** лісистість, прогнозування, супутникові дані, Landsat, нейронні мережі, перцептрон, динаміка вирубки, архітектура мережі, метод навчання, аналіз даних, програмування R.

**Abstract:** The problem of calculating the level of forest cover, including predicting changes in forest cover in a separate forestry, was considered. A comparison of forest cover over 40 years (1984–2024) of the village of Spivakivka, Izyum district, Kharkiv region, was made using the Landsat satellite service. To apply the artificial neural network method, a list of input factors was formed, containing indicators on the selected site in the two previous years and the same indicators on neighboring sites (each of the downloaded images of the selected forestry is divided into 9 squares). A script was created in the programming and data analysis language R, which performs calculations and, allows determining the best architecture of the neural network and a more effective method for its training for a certain data set. A calculation of the dynamics of felling on the selected square is presented (the forecast for the last year provides an error of 9%). After numerous runs of the script, it was found that the best result is provided by a perceptron with one hidden layer with two neurons in it.

**Keywords:** forest cover, forecasting, satellite data, Landsat, neural networks, perceptron, deforestation dynamics, network architecture, training method, data analysis, R programming.

В Україні питання збереження лісів регулюється низкою законодавчих актів [1]. Створення та використання спеціалізованого програмного забезпечення для оцінювання лісистості, оброблення інформації про лісові насадження та виявлення несанкціонованої вирубки лісу, прогнозування зміни лісистості в окремому лісництві може стати суттєвим помічником в цьому напрямку.

Автори розробили програмне забезпечення для розрахунку лісистості та оброблення інформації про лісові насадження [2] на прикладі селища Співаківка в Ізюмському районі Харківської області [3]. Також було зроблено порівняння лісистості за низку років [4–5], для чого було використано ресурс Global Forest Watch (Всесвітня лісова варта) [6] – вебдодаток для моніторингу лісового покриття планети в реальному масштабі часу. Недоліком цього вебдодатку є наявність супутникових знімків тільки за 13 років, до 2013 року.

Автори роботи [7] для дослідження обсягів порушень намету деревостанів в Українських Карпатах упродовж 1984-2016 рр. використали часові ряди зображень, отриманих із супутникових знімків проєкту «Landsat» із застосуванням засобу візуалізації «TimeSync». Автори розрахували середню лісистість для всієї досліджуваної території та



навели перелік основних порушень, що були пов'язані як з антропогенними, так і з природними чинниками.

Таким чином, було прийнято рішення використання ресурсів [8–9] для отримання зображень Придонецького лісництва за 40 років (1984 – 2024) – рис. 1.

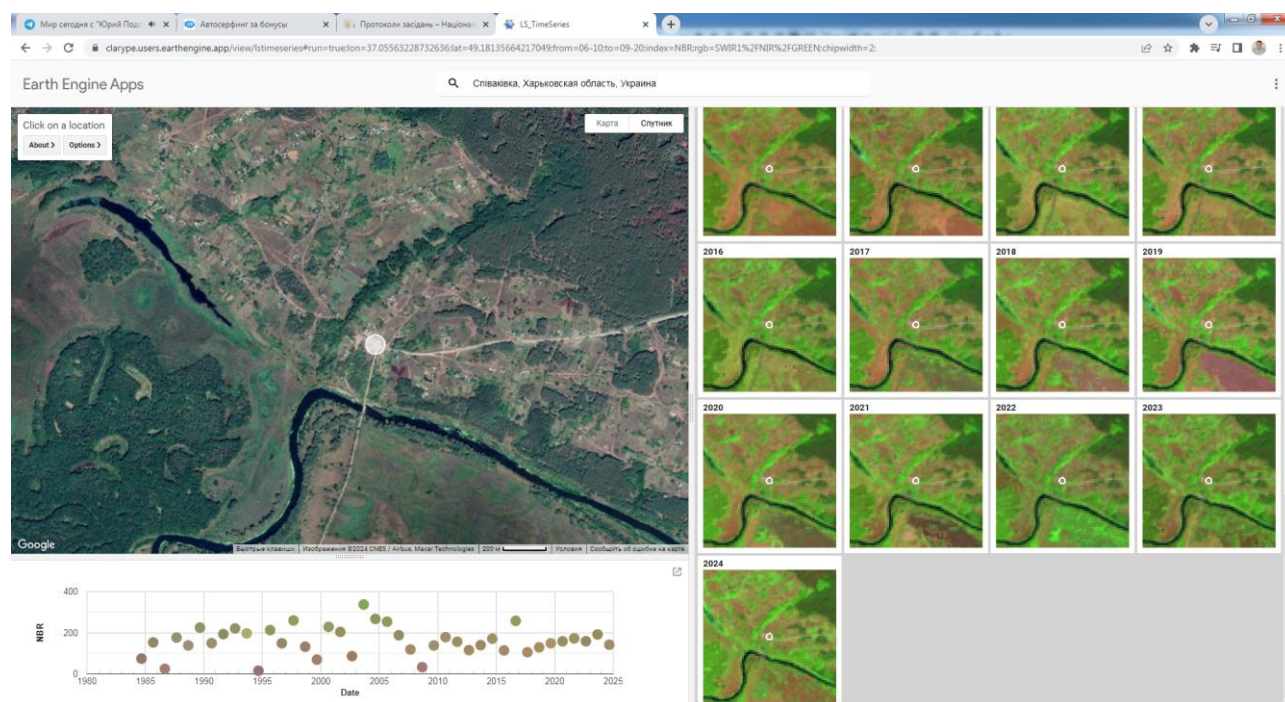


Рисунок 1 – Дані щодо селища Співаківка

У роботі [5] за допомогою сервісу [6] визначався відсоток вирубки, який на сервісі відображався рожевим кольором. Зараз потрібно визначати відсоток зеленого кольору:

$$V = \frac{P_{\text{green}}}{P_z} 100\% \quad (1)$$

де  $V$  – відсоток насиченості;

$P_{\text{green}}$  – кількість пікселів зеленого кольору;

$P_z$  – загальна кількість пікселів на зображенні.

Оскільки нас цікавить не тільки загальна ситуація у лісництві, але і ситуація на окремих ділянках, кожне із завантажених зображень можна розділити на квадрати, а потім аналізувати дані по кожному квадрату. Для кожного квадрату було розраховано насиченість зеленим кольором та збережено у таблиці (рис. 2).

Постає питання прогнозування зміни лісових насаджень на обраній ділянці, тобто зміни відсотка лісистості. Це можна зробити різними шляхами. По-перше, застосувати звичайні рівняння регресії (лінійне, поліноміальне, експоненціальне тощо) окремо до значень кожного квадрату, а також для усього лісництва. Інший шлях – формування переліку вхідних факторів. Як вхідні фактори можна розглядати, по-перше, показники на обраній ділянці у попередні роки (нехай це значення буде дорівнювати двом), а по-друге, ці ж показники на сусідніх



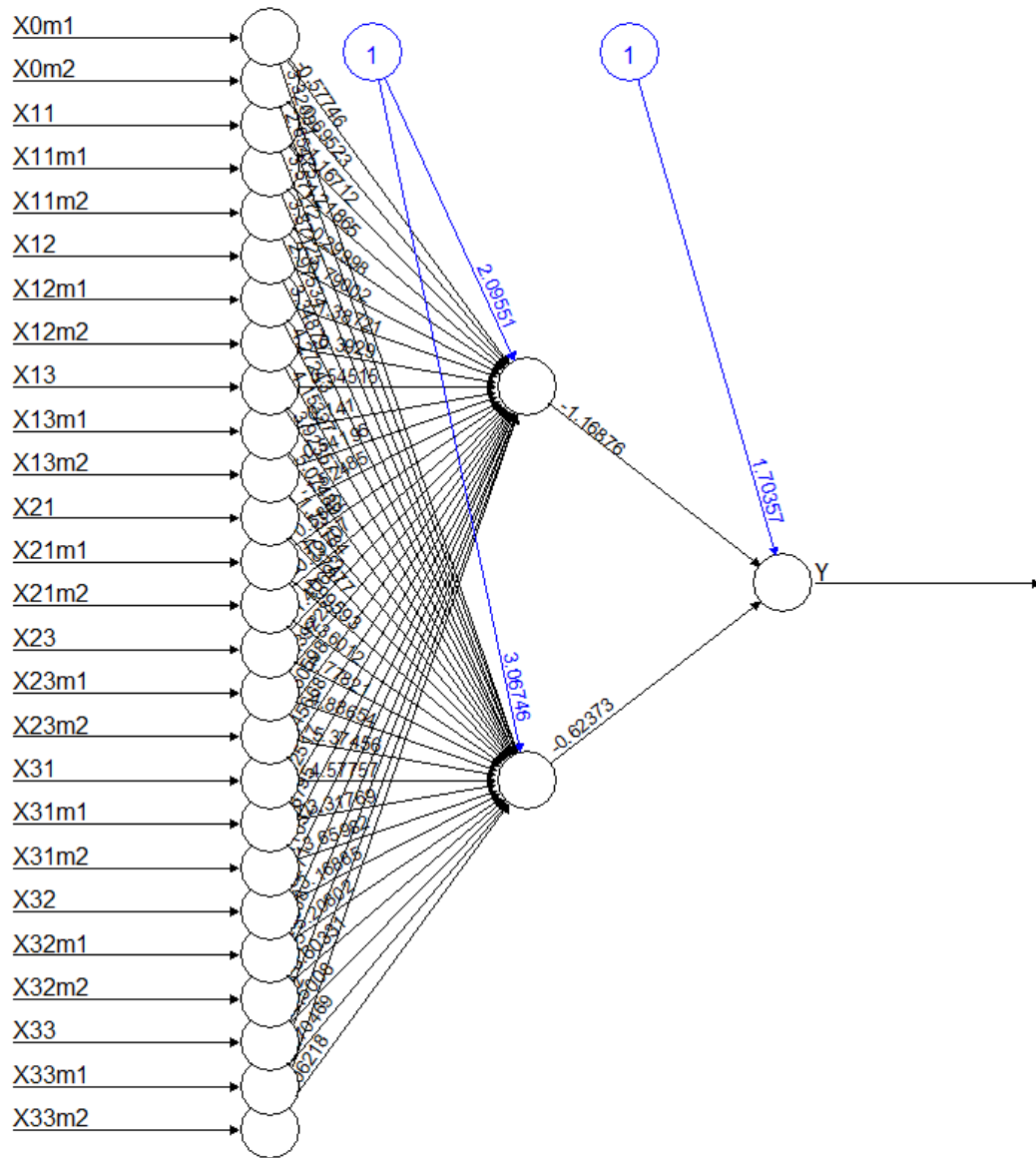


Рисунок 3 – Архітектура нейронної мережі

Таблиця 1 – Результати прогнозування тестової множини

N	Y	Res	error
5	85.57	84.20	0.032065
6	89.87	90.78	0.021238
17	62.98	62.90	0.001956
19	95.64	91.57	0.095026
20	95.64	94.12	0.035584
25	68.72	68.18	0.012687
26	76.55	71.90	0.108633
27	87.16	84.33	0.065978
33	73.50	71.78	0.040224
36	73.65	71.96	0.039438
38	86.21	88.29	0.048462

Таблиця 2 – Результати прогнозування на обраному квадраті MLP 26x2x1

Year	Y	Res	error
2024	76.12	78.19	0.09667

## ВИСНОВКИ

Наведено математичну модель та розрахунки прогнозування зміни рівня лісистості за допомогою супутникового сервісу Landsat і мови програмування та аналізу даних R. Сформовано перелік вхідних факторів, що містять показники на обраній ділянці у два попередні роки та ці ж показники на сусідніх ділянках. Задачу прогнозування розв'язано методом штучних нейронних мереж. Створено скрипт мовою програмування та аналізу даних R, який виконує розрахунки цим методом, а також дозволяє визначити найкращу архітектуру нейронної мережі та більш ефективний метод її навчання для певного набору даних. Наведено розрахунок динаміки вирубки на обраному квадраті (прогноз на останній рік забезпечує похибку в 9%). З'ясовано, що найкращий результат забезпечує перцептрон з одним прихованим шаром з двома нейронами.

## Список посилань.

1. Знищення лісів: причини і наслідки. URL: <https://tvr.biographiya.com/znishhennya-lisiv-prichini-i-naslidki/> (дата звернення: 29.11.2024).
2. Мельников О. Ю., Денисенко В. О. Програмне забезпечення для розрахунку лісистості та оброблення інформації про лісові насадження // Збірник матеріалів III Міжнародної науково-практичної конференції «Використання інформаційних технологій для оптимізації процесів виробництва сільськогосподарської продукції та управління підприємствами» (Агро-IT), (м. Київ, 23-24 березня 2023 р.) / Східноукраїнський національний університет імені Володимира Даля. – Київ, 2023. – С. 39–42.
3. Співаківка (Ізюмський район). Вікіпедія. URL: [https://uk.wikipedia.org/wiki/Співаківка\\_\(Ізюмський\\_район\)](https://uk.wikipedia.org/wiki/Співаківка_(Ізюмський_район)) (дата звернення: 29.11.2024).
4. Денисенко В. О., Мельников О. Ю. Дослідження збільшення вирубки лісу на території Придонецького лісництва // Сучасні комп'ютерні системи та мережі в управлінні: матеріали VI Всеукраїнської наук.-практ. Інтернет-конф. здобувачів вищої освіти та молодих вчених (30 листопада 2023 р., м. Хмельницький, м. Херсон) / за ред. А. А. Григорової. Херсон: Книжкове видавництво ФОП Вишемирський В. С., 2023. С. 161–163.
5. Мельников О. Ю., Денисенко В. О. Прогнозування зміни рівня лісистості за допомогою сервісу Global Forest Watch і мови програмування та аналізу даних R // Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології : зб. наук. пр. / Нац. техн. ун-т «Харків. політехн. ін-т». – Харків: НТУ «ХПІ», 2024. – № 1 (11). – С. 65–69. – DOI: <https://doi.org/10.20998/2079-0023.2024.01.10>
6. Global Forest Watch. URL: <https://www.globalforestwatch.org/> (дата звернення: 29.11.2024).
7. Часковський О. Г., Карабчук Д. Ю., Іванюк А. П. Зміни лісового вкриття українських Карпат за період 1984–2016 років // Науковий вісник НЛТУ України, 2019, т. 29, № 2. – С. 9–13. – DOI: <https://doi.org/10.15421/40290201>
8. OSU eMapR Lab Tools. URL: <https://emapr.ceoas.oregonstate.edu/tools.html> (дата звернення: 29.11.2024).
9. LS\_TimeSeries. URL: <https://clarype.users.earthengine.app/view/lstimeseries#run=true;lon=-122.74036105282948;lat=44.282381579982584;from=06-10;to=09-20;index=NBR;rgb=SWIR1%2FNIR%2FGREEN;chipwidth=2;> (дата звернення: 29.11.2024).
10. Гітіс В. Б. Нейромережні технології: навчальний посібник. – Краматорськ : ДДМА, 2021. – 248 с.
11. Мельников О. Ю. R – мова програмування та аналізу даних: навчальний посібник для здобувачів вищої освіти за спеціальностями «Системний аналіз» та «Інформаційні системи та технології». – Краматорськ: ДДМА, 2023. – 272 с.

## ВИЯВЛЕННЯ ОБ'ЄКТІВ ПРОПАГАНДИ У ТЕКСТОВИХ ПОВІДОМЛЕННЯХ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ ІЗ ВІЗУАЛЬНОЮ ІНТЕРПРЕТАЦІЄЮ РЕЗУЛЬТАТІВ

**Анотація:** Запропоновано метод виявлення об'єктів пропаганди в текстових повідомленнях нейромережевими засобами обробки природної мови із візуальною інтерпретацією результатів. Відмінністю методу є розширення множини об'єктів пропаганди за рахунок додавання варіантів їх словесних подань та використання контекстних вікон для виявлення зв'язків між прийомами та об'єктами пропаганди. Це дозволяє покращити результати виявлення і забезпечити візуальне представлення об'єктів пропаганди, їх словесних подань і важливих зв'язків між ними. Експериментально доведено ефективність підходу, який забезпечує результати, що корелюють з експертними оцінками, і дозволяє візуально спостерігати об'єкти впливу та їх зв'язки в рамках пропагандистських прийомів.

**Ключові слова:** об'єкти пропаганди, прийоми пропаганди, виявлення пропаганди, обробка природної мови

**Abstract:** Proposes method for detecting propaganda objects in text messages using neural network tools for natural language processing with visual interpretation of the results. The difference of the method is the expansion of the set of propaganda objects by adding variants of their verbal representations and using context windows to detect connections between techniques and propaganda objects. This allows to improve the detection results and provide a visual representation of propaganda objects, their verbal representations and important connections between them. The effectiveness of the approach, which provides results that correlate with expert assessments and allows visually observing objects of influence and their connections within the framework of propaganda techniques, has been experimentally proven.

**Keywords:** propaganda objects, propaganda techniques, propaganda detection, natural language processing

### Постановка проблеми

Пропаганда, спрямована на маніпуляцію різними об'єктами для досягнення політичних, соціальних, економічних або культурних цілей, є одним із найбільших викликів сучасності [1]. Об'єктами пропаганди є особи, групи, організації, соціальні верстви, а також явища або інституції, на які спрямовані пропагандистські зусилля з метою впливу на їхню свідомість, емоції, поведінку та суспільну думку. У сучасних умовах важливим завданням є не лише автоматизоване виявлення пропагандистських прийомів, але й визначення об'єктів, на які спрямовані ці прийоми, з візуальною інтерпретацією результатів.

У даній статті представлено метод виявлення об'єктів пропаганди в текстових повідомленнях засобами обробки природної мови. Відмінністю цього методу є розширення множини об'єктів пропаганди завдяки додаванню варіантів їх словесних подань та використанню контекстних вікон для виявлення зв'язків між прийомами та об'єктами пропаганди. Це дозволяє не тільки покращити результати виявлення, але й забезпечити візуальне представлення об'єктів пропаганди, їх словесних подань та важливих зв'язків між ними. Експериментально доведено ефективність підходу, який забезпечує результати, що корелюють з експертними оцінками, і дозволяє візуально спостерігати об'єкти впливу та їх зв'язки в рамках пропагандистських прийомів.

Запропонований підхід корелює із Цілями сталого розвитку ПРООН та сприяє автоматизації процесу виявлення та класифікації пропаганди, забезпечуючи повні, інтерпретовані та зрозумілі результати. Зокрема, застосування методів обробки природної мови для виявлення та класифікації технік і об'єктів пропаганди сприяє досягненню Цілі сталого розвитку ООН №16 шляхом підвищення прозорості інформаційного простору та зміцнення інституційної довіри. Також це підтримує Ціль сталого розвитку ООН №4, розвиваючи медіаграмотність і критичне мислення серед населення, що допомагає ефективно протидіяти дезінформації.

### Аналіз останніх публікацій

Є два основних підходи до ідентифікації пропаганди: через розпізнавання іменованих сутностей (NER) та класифікацію повідомлень [2]. Розглядаючи пропаганду як задачу NER, виникає складність через те, що текстові фрагменти з пропагандистськими елементами

завичай довші, ніж типові об'єкти NER (наприклад, імена чи назви), і можуть складатися з кількох десятків слів. У дослідженні [3] аналізується вплив довжини текстових сегментів на точність виявлення пропаганди, що підтверджує зростання складності із збільшенням довжини діапазонів. Було випробувано кілька популярних методів для цієї задачі, виміряно, наскільки добре вони відображають розподіл довжини текстових фрагментів, а також запропоновано підхід із адаптивним рівнем згортки, який покращує обмін інформацією між віддаленими словами. Це рішення сприяє більш точному відновленню довжини тексту без втрати загальної ефективності.

У рамках досліджень, орієнтованих на виявлення пропаганди на рівні документів, акцент зроблено на оцінці тексту як цілісного елемента і його окремих речень [4]. Для побудови ознак використовуються різні методи: статистичні індикатори, векторизація тексту [5], лінгвістичне маркування, а також розпізнавання тригерів, таких як абсолютні займенники або підсилювальні слова.

Експериментальні результати продемонстрували, що модель, застосована до аналізу на рівні документа, досягла точності 0,943. Вона змогла правильно класифікувати 6097 непропагандистських статей і 694 пропагандистські статті. Підхід, орієнтований на аналіз окремих речень, показав нижчі результати: точність склала 0,744. Він успішно ідентифікував 205 пропагандистських речень і 1917 непропагандистських, проте 731 статтю було класифіковано невірно.

Аналіз пов'язаних робіт у сфері виявлення прийомів та об'єктів пропаганди виявив низку проблем. По-перше, існує відсутність комплексного аналізу взаємозв'язків між техніками та об'єктами пропаганди в текстах. По-друге, бракує узагальнень для об'єктів пропаганди та їх альтернативних згадувань. Пропаганда, яка виявляється тільки через пошук іменованих сутностей, не демонструє спрямованості технік. Також, техніки пропаганди, що виявляються на рівні документу, не відображають об'єктів впливу. При виявленні пропаганди, як завдання пошуку іменованих сутностей, об'єкти часто подаються власними назвами, що охоплює питання «на кого?», однак не охоплює питання «На що?» мають спрямування використані прийоми

### **Мета роботи та постановка завдань**

Мета роботи полягає в створенні методу виявлення об'єктів пропаганди засобами обробки природної мови з візуальною інтерпретацією результатів, який дозволить у пропагандистських повідомленнях знаходити на кого, і на що, спрямовані конкретні використані в повідомленні прийоми пропаганди, а також бачити візуальну інтерпретацію результату.

### **Виклад основного матеріалу**

В рамках підходу до виявлення об'єктів пропаганди у текстових повідомленнях засобами обробки природної мови із візуальною інтерпретацією результатів буде використано множину нейромережових моделей для ідентифікації пропагандистських прийомів для подальшого співвіднесення їх з знайденими об'єктами. Кожна з 17 моделей була попередньо навчена для виявлення кожного з прийомів пропаганди: «Appeal to fear-prejudice», «Causal Oversimplification», «Doubt», «Exaggeration», «Flag-Waving», «Labeling», «Loaded Language», «Minimisation», «Name Calling», «Repetition», «Appeal to Authority», «Black and White Fallacy», «Reductio ad hitlerum», «Red Herring», «Slogans», «Thought terminating Cliches», «Whataboutism» [6] відповідно.

Метод виявлення об'єктів пропаганди у текстових повідомленнях із застосуванням засобів обробки природної мови та візуальної інтерпретації результатів базується на нейромережових моделях глибокого навчання та складається з кількох етапів. Спершу здійснюється ідентифікація об'єктів пропаганди шляхом розпізнавання іменованих сутностей (NER). На цьому етапі проводиться попередня обробка тексту, що включає видалення повторів серед іменованих сутностей на рівні їх лем.

Наступним кроком є розширення множини об'єктів пропаганди за рахунок визначення альтернативних варіантів словесного подання іменованих сутностей. Після цього формуються контекстні вікна для кожного об'єкта пропаганди, з урахуванням заданого порогового

значення мінімального розміру вікна. В межах цих контекстних вікон оцінюється інтенсивність використання прийомів пропаганди за допомогою нейромережових моделей.

На фінальному етапі будується множина важливих зв'язків між об'єктами та прийомами пропаганди, враховуючи порогові значення мінімального рівня прояву пропагандистських прийомів. Отримані результати дозволяють не лише виявити об'єкти та спрямованість пропаганди, але й забезпечити візуалізацію зв'язків між прийомами та їх цільовими об'єктами.

Схематичне представлення запропонованого методу наведено на рисунку 1. Вхідними даними для реалізації методу є текст для аналізу, множина ідентифікованих прийомів пропаганди у тексті та набір попередньо навчених нейромережових моделей, адаптованих для аналізу кожного прийому. Результатом першого етапу є множина об'єктів пропаганди, визначених за допомогою NER, без повторів.

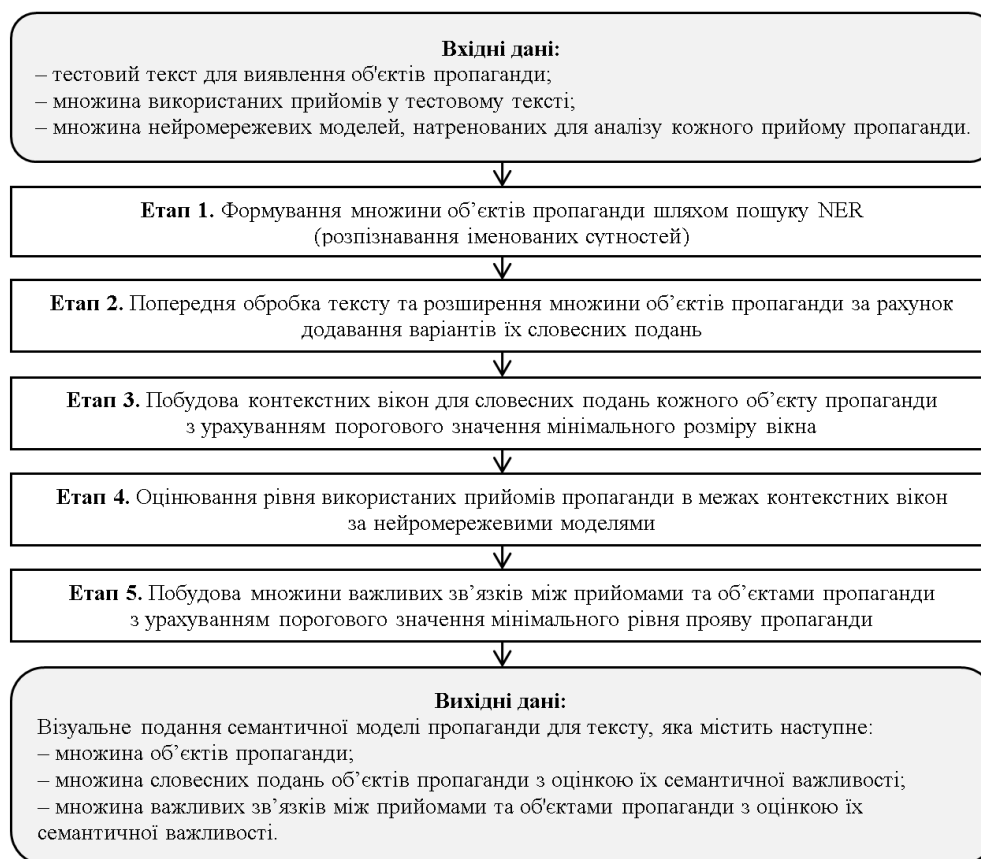


Рисунок 1. Схема методу виявлення об'єктів пропаганди засобами обробки природної мови з візуальною інтерпретацією результатів

На другому етапі методики до кожної ідентифікованої іменованої сутності здійснюється пошук схожих за значенням слів-об'єктів. Це обумовлено тим, що поняття об'єктів пропаганди є ширшим за NER і включає не лише іменовані сутності, але й культурні аспекти, соціальні групи чи узагальнені категорії, об'єднані за спільними характеристиками. Для цього використовується попередньо навчена модель FastText, розроблена Facebook AI Research, яка базується на архітектурах «CBOW» і «Skip-gram». Ця модель дозволяє аналізувати контекст слів, встановлювати семантичні зв'язки та виявляти схожі об'єкти, що дає змогу розширити спектр виявлених об'єктів пропаганди.

У рамках роботи FastText донавчається на текстах, які містять пропаганду, для забезпечення специфічності до задачі. В результаті цього етапу формується розширена множина об'єктів пропаганди, яка включає альтернативні варіанти їх словесних подань.

Мінімальний рівень семантичної близькості визначається емпірично залежно від специфіки завдання; у даному випадку порогове значення не використовувалося.

На третьому етапі формуються контекстні вікна для кожного об'єкта пропаганди. Контекстним вікном вважається речення, у якому згадується конкретний об'єкт. Якщо одне речення містить кілька об'єктів пропаганди, контекстне вікно створюється один раз і включає всі об'єкти. У разі, якщо об'єкт пропаганди має кілька словесних подань, контекстні вікна дублюються для кожного з них, зберігаючи зв'язок із вихідним об'єктом. Мінімальний розмір контекстного вікна визначається пороговим значенням, яке встановлюється відповідно до вимог аналізу.

На четвертому етапі методу здійснюється аналіз контекстних вікон для визначення рівня використання пропагандистських прийомів. Це реалізується через векторизацію текстового контенту контекстних вікон із застосуванням відповідних векторизаторів, після чого нейромережеві моделі аналізують приналежність кожного контекстного вікна до конкретних прийомів пропаганди. Оцінка здійснюється для всіх виявлених у тексті прийомів, що дозволяє визначити, які саме з них були задіяні у межах кожного контексту.

На фінальному, п'ятому етапі, будується множина важливих зв'язків між пропагандистськими прийомами та об'єктами. Це виконується з урахуванням порогового значення мінімального рівня прояву пропаганди. У випадках, коли сила прояву прийому в межах контекстного вікна не перевищує встановлений поріг, такий прийом не вважається застосованим до відповідної групи об'єктів.

Такий підхід забезпечує наочне подання результатів, сприяючи ефективному аналізу повідомлень і розумінню взаємозв'язків між об'єктами та прийомами пропаганди.

Для оцінки ефективності розробленого методу виявлення об'єктів пропаганди було створено спеціалізоване програмне забезпечення, яке дозволяє ідентифікувати об'єкти пропаганди, зіставляти їх із використаними прийомами та відображати результати у формі візуальної аналітики. Отримані дані порівнювалися з висновками авторитетних ресурсів і експертів у сфері протидії пропаганді, що дало змогу оцінити якість запропонованого підходу.

Для тестування використовувалися розмічені повідомлення із соціальних мереж, підготовлені Центром стратегічних комунікацій [7], які містили експертні висновки. Це забезпечило можливість порівняння результатів роботи методу з незалежними оцінками фахівців. Програмне забезпечення для реалізації методу було розроблено у вигляді вебзастосунку на мові програмування Python. В межах створеного програмного забезпечення застосовано: 17 попередньо навчених нейромережевих моделей, створених на основі попередніх досліджень; нейромережеву бібліотеку Stanza для розпізнавання іменованих сутностей (NER); фреймворк Flask для організації вебінтерфейсу; модель FastText, адаптовану до специфіки аналізу пропагандистських повідомлень шляхом донавчання. Приклад повідомлення та його аналізу авторитетним джерелом наведено на рисунку 2.

У ході дослідження ефективності запропонованого методу виявлення об'єктів і прийомів пропаганди було встановлено, що результати, отримані за допомогою розробленого підходу, демонструють високу кореляцію з експертними оцінками, представленими Центром стратегічних комунікацій [18].



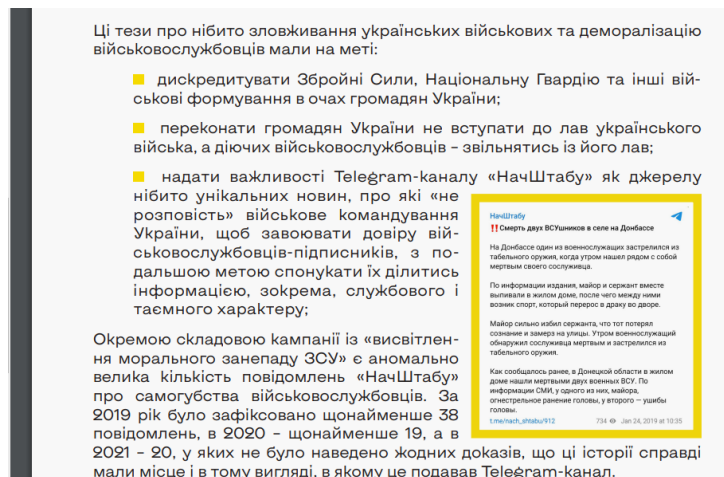


Рисунок 2. Аналіз повідомлення що містить пропаганду від авторитетного джерела [18]

Для підтвердження наведено приклад аналізу допису з пропагандистського каналу (рисунок 2). У цьому прикладі результати автоматичного аналізу, виконаного за допомогою розробленого методу, співпали з висновками експертів. Це свідчить про здатність методу точно ідентифікувати об'єкти пропаганди та визначати, які прийоми було використано для їх маніпулятивного впливу.

Отримані результати підтверджують практичну застосовність розробленого підходу для автоматизованого аналізу пропагандистських повідомлень і можливість його використання в реальних умовах для підтримки роботи аналітиків та дослідників. Результати наведені на рисунку 3.

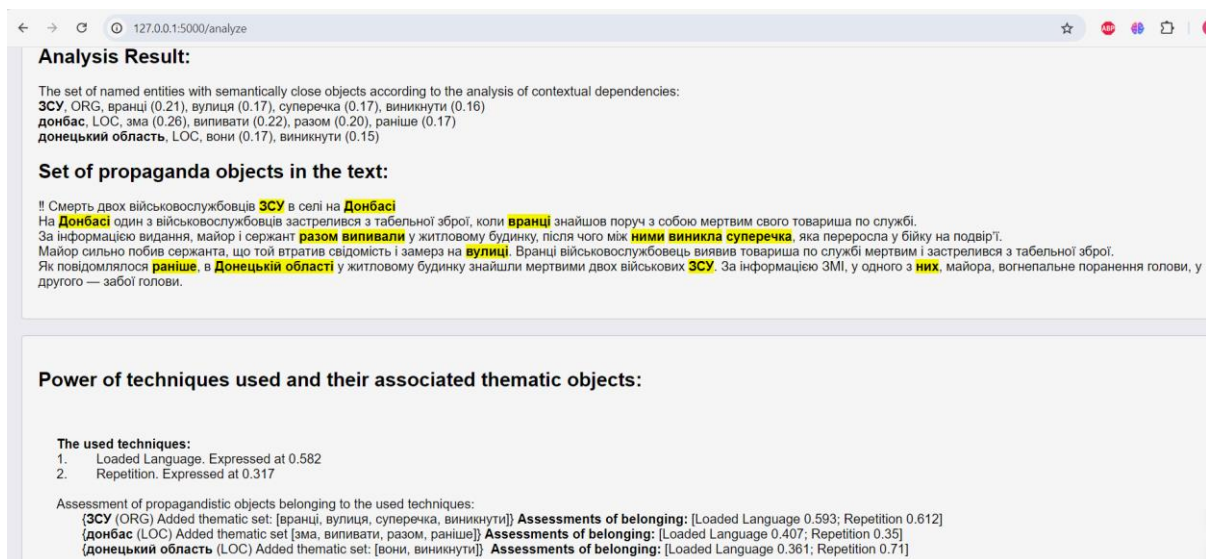


Рисунок 3. Візуальна інтерпретацією прийнятих рішень за методом виявлення об'єктів пропаганди

Аналіз за допомогою розробленого програмного забезпечення показав використання пропагандистських прийомів «Loaded Language» (0.582) і «Repetition» (0.317), ідентифікував об'єкти пропаганди (наприклад, ЗСУ, Донбас, Донецька область) разом із семантично близькими словами та оцінив відповідність об'єктів пропаганди до прийомів: ЗСУ («Loaded Language» – 0.593, «Repetition» – 0.612), Донбас («Loaded Language» – 0.407, «Repetition» – 0.35), Донецька область («Loaded Language» – 0.361, «Repetition» – 0.71), з подальшим візуальним відображенням знайдених об'єктів у тексті.

У результаті дослідження запропонованого методу виявлено, що він дозволяє отримувати результати, які корелюють із результатами з авторитетних маркованих джерел. За допомогою застосування комплексного підходу до виявлення пропаганди, та з використанням візуальної інтерпретації результатів, вирішується задача взаємозв'язків між використаними прийомами та об'єктами пропаганди.

## ВИСНОВКИ.

Розроблено метод виявлення об'єктів пропаганди засобами обробки природної мови з візуальною інтерпретацією прийнятих рішень, що відрізняється від існуючих розширенням множини об'єктів пропаганди завдяки додаванню варіантів їх словесних подань і використанню контекстних вікон для виявлення взаємозв'язків між використаними прийомами та об'єктами пропаганди. Це покращує результати виявлення та їх візуальне представлення. Метод включає розпізнавання іменованих сутностей, попередню обробку тексту, розширення множини об'єктів пропаганди, побудову контекстних вікон, оцінювання рівня використаних прийомів та побудову важливих зв'язків між прийомами та об'єктами пропаганди.

Для підвищення точності та якості виявлення прийомів та об'єктів пропаганди за семантичними маркерами у повідомленнях засобами обробки природної мови з подальшою інтерпретацією результатів, було розроблено підхід, який дозволяє ідентифікувати об'єкти пропаганди у текстах, а також на кого і на що спрямовані пропагандистські прийоми. Метод вирішує проблеми відсутності комплексного аналізу взаємозв'язків прийомів та об'єктів пропаганди в повідомленнях і відсутності узагальнень для об'єктів пропаганди та їх альтернативних згадок. Експериментально доведено ефективність підходу, що дозволяє, окрім пошуку NER за допомогою бібліотеки нейронної мережі «STANZA», розширювати перелік об'єктів пропаганди за допомогою бібліотеки машинного навчання «FastText», а також оцінювати їх зв'язок з використаними прийомами. Результати методу корелюють з експертними оцінками, а візуальна аналітика забезпечує наочне спостереження об'єктів впливу в рамках пропагандистських прийомів.

## Список посилань:

1. M.B. Shevtsiv, K. A. Honcharuk, *Propaganda as socio-legal phenomenon: problems of understanding*, in: *Current problems of historical and legal and of international legal science. South Ukrainian magazine 1*, 2019, pp. 119 – 122
2. Молчанова М.О. Нейромережеве виявлення і класифікація прийомів та об'єктів пропаганди у текстовому контенті. *Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах»*, № 4, 2024. с. 153-161. <https://doi.org/10.31891/2219-9365-2024-80-19> URL: <https://vottp.khmnpu.edu.ua/index.php/vottp/article/view/385/361>
3. Przybyla P. *Long Named Entity Recognition for Propaganda Detection and Beyond* / P. Przybyla, K. Kaczynski // *Proceedings of the International Conference of the Spanish Society for Natural Language Processing*. – 2023
4. Vysotska V. *Information technology for recognizing propaganda, fakes and disinformation in textual content based on nlp and machine learning methods*. *Radio Electronics, Computer Science, Control*. 2024. No. 2. P. 126. URL: <https://doi.org/10.15588/1607-3274-2024-2-13> (date of access: 29.11.2024).
5. *From Bag-of-Words to Transformers: A Comparative Study for Text Classification in Healthcare Discussions in Social Media* / E. De Santis et al. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2024. P. 1–15. URL: <https://doi.org/10.1109/tetci.2024.3423444> (date of access: 29.11.2024).
6. Krak I., Molchanova M., Mazurets O., Sobko O., Zalutska O., Barmak O. *Method for Neural Network Detecting Propaganda Techniques by Markers With Visual Analytic*. *CEUR Workshop Proceedings*, 2024, vol. 3790, pp. 158-170.
7. *Tsentr stratehichnykh komunikatsii [Elektronnyi resurs]*. – Rezhym dostupu: <https://spravdi.gov.ua/>

Павлюк О.М., Міщук М.В., Заболотна А.О., Літовська О.В., Костінський Д.О.,  
(Національний університет «Львівська політехніка», м. Львів, Україна)

## НЕЙРОМЕРЕЖНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ КОМПЛЕКСНОЇ ДІЯЛЬНОСТІ ПРОМИСЛОВОГО ПЕРСОНАЛУ ЗА ДАНИМИ СМАРТ- ГОДИННИКІВ

**Анотвція:** Сучасні підходи до розпізнавання комплексної людської діяльності в реальному часі відкривають нові можливості для їх використання в Індустрії 5.0. За рахунок цього можна покращити людино-машинну взаємодію та оцінку продуктивності працівників. У роботі представлено систему для розпізнавання комплексної діяльності промислового персоналу у внутрішніх логістичних системах з використанням смарт-годинників. Розроблено аплікацію за допомогою якої зібрано унікальний набір даних. Застосовано виявлення аутлаєрів у зібраних даних та згладжування шумів за допомогою фільтрів. Для класифікації простих активностей використано нейронну мережу DenseNet121, яка досягла F1-оцінки 91,01%, для комплексних – 87,44% зі стратегією спільних даних.

**Ключові слова:** смарт-годинник; промисловий персонал; базовий класифікатор; комплексна діяльність; класифікація; передбачення.

**Abstract:** Modern approaches to recognizing complex human activity in real time open up new opportunities for their use in Industry 5.0. This can improve human-machine interaction and employee productivity. The paper presents a system for recognizing complex activities of industrial personnel in internal logistics systems using smart watches. An application has been developed with the help of which a unique data set has been collected. Outlier detection in the collected data and noise smoothing using filters were applied. For the classification of simple activities, the DenseNet121 neural network was used, which achieved an F1 score of 91.01%, for complex activities – 87.44% with a shared data strategy.

**Keywords:** smart watch; industrial personnel; basic classifier; complex activity; classification; prediction.

### Постановка проблеми

Використання портативних технологій для розпізнавання людської діяльності (РЛД) набуло активного розвитку і знайшло застосування в: медицині (детекція падінь, епілепсії та моніторингу активності [1, 2]); безпеці (виявлення аномалій [3]); спорті (аналіз тренувань і витрати калорій [4, 5]). З переходом до Індустрії 5.0 відкрилися нові горизонти використання РЛД для покращення людино-машинної взаємодії у системи інтелектуального управління підприємствами [6-8]. Ключовим елементом для досліджень в таких системах є дані зібрані за допомогою сенсорів як з технічних систем так і з промислового персоналу.

Федеративне навчання (ФН), як сучасний метод машинного навчання, дозволяє здійснювати безпечний збір та аналіз інформації не лише з технологічного обладнання, але і з носимих пристроїв промислового персоналу таких як смарт-годинники без передачі до центрального сервера. Це особливо важливо в умовах промислової екосистеми України, де питання конфіденційності та захисту даних є критичними. ФН дозволяє працювати навіть в умовах нестабільної інфраструктури з частими перебоями зв'язку в умовах війни із забезпеченням автоматизації прийняття рішень на основі локально зібраних даних в реальному часі. Це можливо здійснити за рахунок: використання локальних буферів, асиметричного графіку синхронізації, використання Edge-серверів, стійких протоколів передачі даних, локальної автономності, Dropout-резистентності, стиснення та обмеження обсягу даних. Такі система можна легко адаптувати до потреб конкретних українських виробництв, сприяючи їх конкурентоспроможності на міжнародному ринку.

### Аналіз останніх публікацій

Згідно проведеного аналізу літературних джерел, загалом, РЛД можна розділити на 2 типи завдань. Проста повторювана поведінка (біг, сидіння) [9-12] та складні контекстуальні дії (робота, водіння) [13-17] які вимагають класифікації, складних моделей та роботи в режимі

реального часу. Проста поведінка вже ефективно розпізнається, але основна увага в цьому дослідженні полягає в прогнозуванні складних дій, що мають практичне значення в сучасній інтелектуальній індустрії.

Більшість існуючих рішень для моніторингу поведінки промислових працівників базуються на аналізі зображень з камер та отриманих із портативних сенсорів або їх комбінації. Однак підхід, заснований на зображеннях, має свої недоліки, оскільки накладає обмеження на траєкторії руху промислового персоналу, який має знаходитися в зоні видимості камер. Це є неможливим на підприємствах, оскільки працівник повинен виконувати різні дії та одночасно співпрацювати з декількома автономними транспортними засобами (АКТЗ) чи іншими промисловими технічними об'єктами. Тому дослідження зосереджено на розробці підходів до комплексної класифікації та прогнозуванні діяльності промислового персоналу з використанням сенсорів вбудованих у смарт-годинники.

В контексті смарт-індустрії у [16] представлено алгоритм XGBoost показав найкращі результати для класифікації діяльності в логістичних системах. Рішення на основі відеокамер і портативних сенсорів [17–19] потребують значних ресурсів і обмежують мобільність працівників. Альтернативний підхід у [20] використовує сенсори ємності тіла та IMU, але вимагає спеціалізованого обладнання.

### **Мета роботи та постановка завдань**

**Мета роботи:** розробка нейромережної системи аналізу комплексної діяльності промислового персоналу в реальному часі на основі даних з сенсорів смарт-годинників. Для збору даних необхідно розробити аплікацію, яка може обробляти інформацію на пристрої або передавати її на крайовий сервер.

#### **Постановка завдань:**

1. Розробити систему збору даних зі сенсорів смарт-годинників для моніторингу діяльності персоналу в реальному часі.
2. Зібрати дані про типову активність персоналу на підприємствах які працюють з АКТЗ.
3. Застосувати методи виявлення та усунення викидів та шумів у даних акселерометра та гіроскопа.
4. Підготувати дані для тренування моделей машинного навчання.
5. Розробити алгоритми для класифікації комплексної діяльності, що підтримує розподілені обчислення та ФН та перевірити його ефективність.

#### **Система збору інформації за допомогою сенсорів смарт-годинника.**

Розроблено розподілену систему збору та аналізу даних яка складається з розумного годинника Samsung Galaxy Watch 5, додатку для WearOS та хмарного сервера. Додаток написаний мовою Kotlin для операційної системи WearOS передбачає збір даних за допомогою сенсорів смарт-годинника; контроль процесу експерименту через інтерфейс; передчу даних у хмару. Хмарний сервер побудований з використанням MySQL-Server на платформі PaaS. Аплікація дозволяє одночасний збір даних від кількох суб'єктів промислового персоналу. На рис. 1 проілюстровано загальну структуру розробленої системи по збору даних.

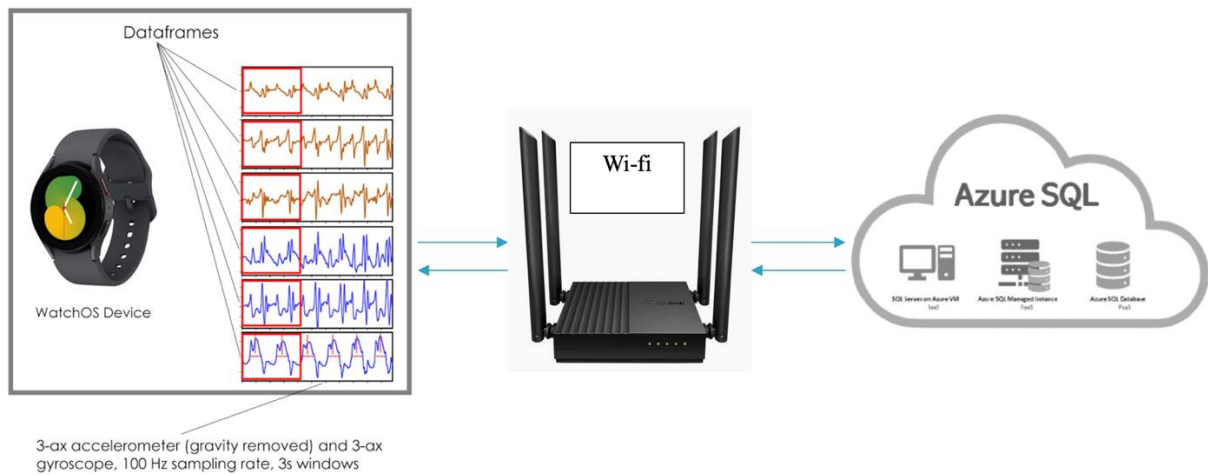


Рис.1. Структура розробленої системи збору даних.

За допомогою розробленої програмної аплікації для смарт-годинника зібрано дані з триосових акселерометра та гіроскопа (шість каналів) з частотою 100 Гц для п'яти представників промислового персоналу по двох агрегатах, загалом 3,28 години шестиканальних даних. Кожен 3-секундний фрагмент даних разом з інформацією (яка рука, мітки часу, ідентифікатор суб'єкта) стискається та надсилається як фрейм даних. Фрагменти менше трьох секунд відкидаються. Кадри з однаковими мітками часу об'єднуються в групи, а неповні групи відкидаються.

### Методологія комплексного РЛД

Запропонована методологія використовує стекінг класифікаторів для розпізнавання та прогнозування типу агрегату на основі сигналів сенсорів. Базовий класифікатор навчається на 3-секундних 6-канальних даних, оброблених НВП, для розпізнавання базових активностей, а метакласифікатор класифікує агрегати на основі послідовностей результатів базового класифікатора, з використанням трансферного навчання для покращення продуктивності і уникнення перенавчання.

На першому етапі дані обробляються для усунення аутлаєрів та згладжування шумів, створюючи неперервні 6-канальні послідовності сигналів, що відображають виконання агрегату. На другому етапі застосовується неперервне вейвлет перетворення (НВП) до кожного каналу сигналів. Третій етап включає тренування базового класифікатора для розпізнавання базових активностей, після чого мітки для них перевизначаються. На четвертому етапі метакласифікатор та предиктор навчаються на послідовностях результатів базового класифікатора для розпізнавання комплексної діяльності активності.

Викиди в даних можуть знизити точність системи розпізнавання комплексної діяльності, тому було застосовано різні методи виявлення аутлаєрів. На рис 2. показані серії даних з викидами, визначеними різними методами: стандартного відхилення (STD), логарифмічного стандартного відхилення (Log-STD), просторової кластеризації додатків із шумом на основі щільності (DBSCAN), міжвертикального розмаху (IQR). Червоні точки показують викиди для кожного методу.

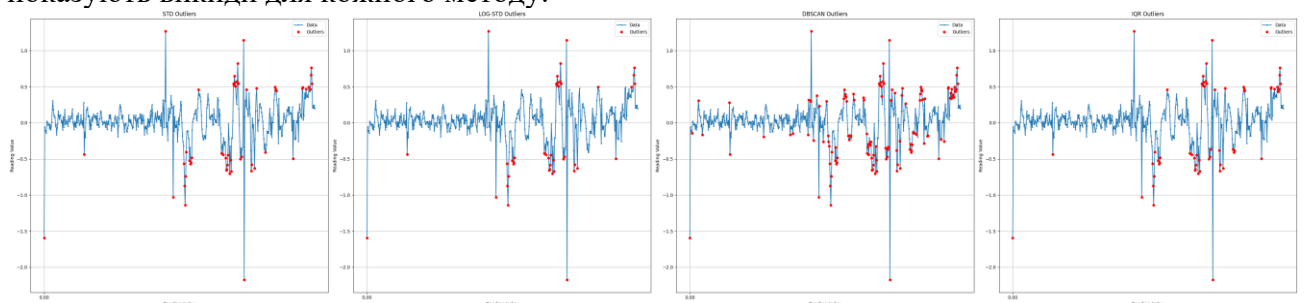


Рис. 2. Виявлення викидів за допомогою методів STD, Log-STD, DBSCAN, IQR.

Як видно, що різні методи виявляють різну кількість викидів. DBSCAN виявляє найбільшу кількість викидів, що може свідчити про його чутливість до аномалій у даних. Для мінімізації впливу шумів використовуються методи фільтрації, зокрема ковзне середнє, експоненціальне згладжування та фільтр Савіцького-Голея (див. рис.3)

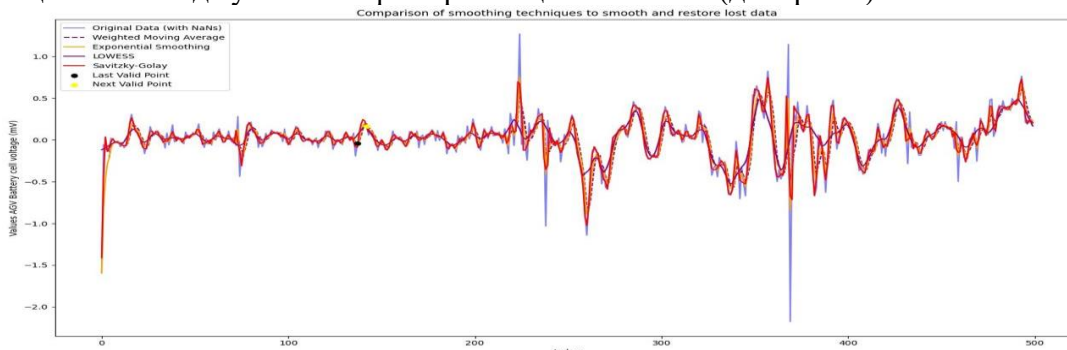


Рис. 3 Порівняння методів згладження за допомогою фільтрів

Як видно по результатах рис. 3. шуми доцільно згладжувати за допомогою локальної регресії. Проте для її реалізації необхідна достатня кількість даних у кожному інтервалі.

Вихідні дані за допомогою неперервне вейвлет перетворення (НВП) можна представити як двовимірну теплову карту (скалограму). Вона дозволяє перевести класифікацію часових рядів у задачу класифікації зображень. Тому далі можна використовувати наприклад згорткові нейронні мережі (ЗНМ). Для цієї роботи обрано материнський вейвлет Морле з параметрами  $a$  від 0 до 256 і  $b$  від 0 до 300. Такі параметри підтверджені попередніми дослідженнями для задач РЛД на основі портативних сенсорів [3, 12].

Попереднього опрацювання даних для нейронних мереж здійснюють у шість етапів. Перший етап включає кадри поза експериментом. Другий виявляє та виділяє розриви між кадрами. Третій зводить послідовності до 20 кадрів (60 секунд). Четвертий виконує 50% перекриття для розширення набору. П'ятий розділяє набір на підмножини для навчання, тестування та валідації. Для розділення даних використовуються дві стратегії. Перша забезпечує унікальні дані для кожного класифікатора, а друга — більше даних для мета-класифікатора. Шостий етап включає застосування НВП з материнським вейвлетом Морле та параметром масштабування від 0 до 256.

Як базовий класифікатор використано модель DenseNet121, спеціально адаптовану для задач РЛД, попередньо навчена на наборі KU-HAR [12, 21] з F1-оцінкою 97,52%. Модель адаптовано, замінивши верхній шар на новий із чотирма нейронами (ініціалізація Xavier). Заморожено 136 шарів для роботи з малими наборами даних. Навчання проводилося 100 епох з оптимізатором Adam, батчем 32 та зворотними викликами для контролю. Результати наведено для найкращого з 10 запусків.

Далі порівняно мета-класифікатори на базі архітектур LSTM, BiLSTM, GRU, BiGRU та ЗНМ. Найкращий результат показала ЗНМ на підмножині тестів у табл.1

Таблиця 1. Результати класифікації базового класифікатора на підмножині тестів.

Accuracy	Precision	Recall	AUC	F1-score
90.90%	91.33%	90.69%	97.26%	91.01%

Після навчання та оцінки базового класифікатора мітки набору даних для всіх зразків було оновлено на основі результатів від нейронів верхнього повністю підключеного шару навченого базового класифікатора. Далі застосовано дві стратегії для класифікації агрегатів. При першій стратегії розділення 40% набору даних використовується для навчання базового класифікатора та валідації мета-класифікатора, інші 40% для навчання мета-класифікатора та валідації базового класифікатора, та останні 20% для тестування. Ця стратегія забезпечує унікальні дані для навчання моделей кожного рівня, що є найкращим сценарієм, але яка

потенційно надає недостатньо даних для навчання мета-класифікатора, оскільки він не використовує тансферне навчання. Друга стратегія виділяє 40% на навчання обох класифікаторів, ще 40% на валідацію базового класифікатора та донавчання метакласифікатора і решту 20% на тестування. Друга стратегія надає більше даних для мета-класифікатора, але може призвести до того, що він не зафіксує помилки базового класифікатора на нових даних.

Таблиця 2. Результати класифікації мета-класифікаторів на підмножині тестів першою стратегією

Classifier	Accuracy	Precision	Recall	AUC	F1-score
CNN	79.17%	79.01%	79.17%	84.34%	79.07%
Single-layer LSTM	75.00%	76.11%	75.00%	78.03%	75.32%
Multi-layer LSTM	76.39%	77.15%	76.39%	77.41%	76.63%
Single-layer BiLSTM	77.78%	78.84%	77.78%	79.13%	78.07%
Multi-layer BiLSTM	75.00%	75.49%	75.00%	79.91%	75.19%
Single-layer GRU	76.39%	76.21%	76.39%	84.51%	76.28%
Multi-layer GRU	75.00%	75.00%	75.00%	79.92%	75.00%
Single layer BiGRU	75.00%	74.45%	75.00%	78.13%	74.48%
Multi-layer BiGRU	73.61%	74.42%	73.61%	79.68%	73.89%

Модель на основі CNN досягла F1-оцінки 79,07%, а одношарова мережа BiLSTM — 73,89%. Інші моделі показали F1 близько 76%. Друга стратегія поділу наведена в таблиці 3.

Таблиця 3. Результати класифікації мета-класифікаторів на підмножині тестів другою стратегією.

Classifier	Accuracy	Precision	Recall	AUC	F1-score
CNN	87.50%	87.43%	87.50%	92.40%	87.44%
Single-layer LSTM	69.44%	71.42%	69.44%	76.93%	69.96%
Multi-layer LSTM	72.22%	72.75%	72.22%	78.70%	72.43%
Single-layer BiLSTM	72.22%	75.04%	72.22%	78.43%	72.75%
Multi-layer BiLSTM	73.61%	74.42%	73.61%	78.76%	73.89%
Single-layer GRU	72.22%	76.08%	72.22%	79.74%	72.77%
Multi-layer GRU	72.22%	72.22%	72.22%	80.84%	72.22%
Single-layer BiGRU	72.22%	75.04%	72.22%	77.93%	72.75%
Multi-layer BiGRU	73.61%	76.92%	73.61%	80.84%	74.13%

Друга стратегія поділу показала більшу ефективність для CNN з F1 87,44%, в порівнянні з першою стратегією. Моделі LSTM, BiLSTM і BiGRU також показали кращі результати з додатковими шарами.

## ВИСНОВКИ

У цьому дослідженні розроблено багатоетапний підхід у режимі реального часу для комплексної класифікації активності промислового персоналу на підприємствах які використовують АКТЗ. Він використовує смарт-годинники, стекування класифікаторів, вейвлет-перетворень і тансферне навчання. Розроблено аплікацію для збору даних на основі смарт-годинників. Зібрано унікальний набір даних від п'яти представників промислового персоналу. Як базовий класифікатор використано попередньо навчену модель DenseNet121, яка досягла F1-оцінки 91,01%. Для багатоетапної класифікації порівняли метакласифікатори на основі CNN, LSTM, BiLSTM, GRU і BiGRU. Найкращі результати показала модель на основі CNN з використанням спільних даних для навчання, отримавши F1-оцінку 87,44%.

У майбутньому планується розширення даних і використати гібридні моделі (CNN-LSTM, CNN-GRU) для покращення точності. Такий підхід легко інтегрується та масштабується в інтелектуальні системи управління для оптимізації виробничих процесів.

## Список посилань.

1. Li, H., Shrestha, A., Heidari, H., Le Kernec, J., & Fioranelli, F. (2020). Bi-LSTM Network for Multimodal Continuous Human Activity Recognition and Fall Detection. *IEEE Sensors Journal*, 20(3), 1191–1201. <https://doi.org/10.1109/JSEN.2019.2946095>.
2. Butt, F. S., La Blunda, L., Wagner, M. F., Schäfer, J., Medina-Bulo, I., & Gómez-Ullate, D. (2021). Fall Detection from Electrocardiogram (ECG) Signals and Classification by Deep Transfer Learning. *Information*, 12(2), Article 2. <https://doi.org/10.3390/info12020063>.

3. Dhiman, C., & Vishwakarma, D. K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition. *Engineering Applications of Artificial Intelligence*, 77, 21–45. <https://doi.org/10.1016/j.engappai.2018.08.014>.
4. Nadeem, A., Jalal, A., & Kim, K. (2021). Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model. *Multimedia Tools and Applications*, 80(14), 21465–21498. <https://doi.org/10.1007/s11042-021-10687-5>.
5. Kalpesh, J., Rushikesh, J., Swaraj, K., Rohan, K., & Bharadwaj, R. (2023). Human Physical Activities Based Calorie Burn Calculator Using LSTM. In J. Hemanth, D. Pelusi, & J. I.-Z. Chen (Eds.), *Intelligent Cyber Physical Systems and Internet of Things* (pp. 405–424). Springer International Publishing. [https://doi.org/10.1007/978-3-031-18497-0\\_31](https://doi.org/10.1007/978-3-031-18497-0_31).
6. Mohsen, S., Elkaseer, A., & Scholz, S. G. (2021). Industry 4.0-Oriented Deep Learning Models for Human Activity Recognition. *IEEE Access*, 9, 150508–150521. <https://doi.org/10.1109/ACCESS.2021.3125733>.
7. Niemann, F., Lüdtkke, S., Bartelt, C., & ten Hompel, M. (2022). Context-Aware Human Activity Recognition in Industrial Processes. *Sensors*, 22(1), Article 1. <https://doi.org/10.3390/s22010134>.
8. Cupek, R., et al. (2020). Autonomous Guided Vehicles for Smart Industries – The State-of-the-Art and Research Challenges. In V. V. Krzhizhanovskaya et al. (Eds.), *Computational Science – ICCS 2020* (pp. 330–343). Springer International Publishing. [https://doi.org/10.1007/978-3-030-50426-7\\_25](https://doi.org/10.1007/978-3-030-50426-7_25).
9. Fang, L., Yishui, S., & Wei, C. (2016). Up and Down Buses Activity Recognition Using Smartphone Accelerometer. In 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference (pp. 761–765). IEEE. <https://doi.org/10.1109/ITNEC.2016.7560464>.
10. Khatun, M. A., & Yousuf, M. A. (2020). Human Activity Recognition Using Smartphone Sensor Based on Selective Classifiers. In 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1–6). IEEE. <https://doi.org/10.1109/STI50764.2020.9350486>.
11. Gaud, N., Rathore, M., & Suman, U. (2024). Hybrid Deep Learning-Based Human Activity Recognition (HAR) Using Wearable Sensors: An Edge Computing Approach. In A. Swaroop et al. (Eds.), *Proceedings of Data Analytics and Management* (pp. 399–410). Springer Nature. [https://doi.org/10.1007/978-981-99-6544-1\\_30](https://doi.org/10.1007/978-981-99-6544-1_30).
12. Pavliuk, O., Mishchuk, M., & Strauss, C. (2023). Transfer Learning Approach for Human Activity Recognition Based on Continuous Wavelet Transform. *Algorithms*, 16(2), Article 2. <https://doi.org/10.3390/a16020077>.
13. Rosen, E., & Senkal, D. (2022). CHARM: A Hierarchical Deep Learning Model for Classification of Complex Human Activities Using Motion Sensors. arXiv. <https://doi.org/10.48550/arXiv.2207.07806>.
14. Liu, Y., Nie, L., Han, L., Zhang, L., & Rosenblum, D. S. (2016). Action2Activity: Recognizing Complex Activities from Sensor Data. arXiv. <https://doi.org/10.48550/arXiv.1611.01872>.
15. Ryoo, M. S., & Aggarwal, J. K. (2006). Recognition of Composite Human Activities Through Context-Free Grammar Based Representation. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (pp. 1709–1718). IEEE. <https://doi.org/10.1109/CVPR.2006.242>.
16. Ding, G., & Yao, A. (2022). Temporal Action Segmentation with High-Level Complex Activity Labels. arXiv. <https://doi.org/10.48550/arXiv.2108.06706>.
17. Syed, A. S., Syed, Z. S., Shah, M. S., & Saddar, S. (2020). Using Wearable Sensors for Human Activity Recognition in Logistics: A Comparison of Different Feature Sets and Machine Learning Algorithms. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(9). <https://doi.org/10.14569/IJACSA.2020.0110976>.
18. Tammvee, M., & Anbarjafari, G. (2021). Human Activity Recognition-Based Path Planning for Autonomous Vehicles. *SIViP*, 15(4), 809–816. <https://doi.org/10.1007/s11760-020-01800-6>.
19. Al-Amin, Md., et al. (2019). Action Recognition in Manufacturing Assembly Using Multimodal Sensor Fusion. *Procedia Manufacturing*, 39, 158–167. <https://doi.org/10.1016/j.promfg.2020.01.288>.
20. Tao, W., Leu, M. C., & Yin, Z. (2020). Multi-Modal Recognition of Worker Activity for Human-Centered Intelligent Manufacturing. *Engineering Applications of Artificial Intelligence*, 95, 103868. <https://doi.org/10.1016/j.engappai.2020.103868>.
21. Pavliuk, O., & Mishchuk, M. (2022). A Novel Deep-Learning Model for Human Activity Recognition Based on Continuous Wavelet Transform. *CEUR Workshop Proceedings*, 236–245.



Петров Д. Д., (Національний університет «Львівська політехніка», м. Львів, Україна)

## ПОКРАЩЕННЯ ЯКОСТІ ЗВ'ЯЗКУ GSM МЕРЕЖІ RL АГЕНТОМ

**Анотація:** Глобальна система мобільного зв'язку (GSM) формується базовими станціями, що надають абонентам канали мережі для комунікації. Зазвичай, якість каналів комунікації налаштовують базової станції налаштовують один раз на весь життєвий цикл, що є причиною неоптимального поділу ресурсу, і як результат гіршої якості зв'язку. У цій статті, пропонується використання агента навчання з підкріпленням (RL) для динамічного підбору оптимальних параметрів базової станції. Запропонований підхід покращує середню якість та стабільність зв'язку із незначним впливом на мережу.

**Ключові слова:** GSM, Глобальна система мобільного зв'язку, навчання з підкріпленням, глибоке навчання з підкріпленням, машинне навчання.

**Abstract:** The Global System for Mobile Communications (GSM) is formed by base stations that provide users with channels for communication. Typically, the quality of communication channels is adjusted once per lifetime of the base station, which causes suboptimal resource allocation and, as a result, inferior communication quality. In this paper, we propose a reinforcement learning (RL) agent application to dynamically select optimal base station parameters. The proposed approach improves the average quality and stability of communication with little impact on the network.

**Keywords:** GSM, Global System for Mobile Communications, RL, DRL, ML.

### Вступ

Не зважаючи на розвиток технології мобільного зв'язку, мережі другого покоління (2G) продовжують відігравати важливу роль у забезпеченні базових комунікаційних потреб. Особливо в часи дефіциту електроенергії такі мережі особливо актуальні завдяки своїй простоті й енергоефективності. Однак оптимізація ресурсів у таких мережах є складним завданням через обмежену кількість частотного спектра, мінливі умови навантаження та наявність інтерференції.

Поточні підходи до розподілу ресурсами часто базуються на фіксованих алгоритмах, які не можуть адаптуватися до динамічних змін середовища. Натомість методи навчання з підкріпленням (Reinforcement Learning, RL) задні пошуку оптимальних стратегій на основі накопиченого досвіду та в автономному режимі змінювати параметри мережі для забезпечення кращої якості.

Ціль цієї статті в розробці алгоритму, що здатен покращити якість зв'язку базових станцій динамічно модифікуючи їх параметри. Для досягнення цієї цілі будуть використані різні методи, такі як методи прогнозування, навчання з підкріпленням та симуляції середовища.

Обґрунтування вибору цієї теми полягає в тому, що поточні підходи до підбору параметрів вимагають значних людського ресурсу, що робить динамічну параметризацією практично неможливою. Також поки сучасні дослідження зосереджені на оптимізації air-ground та space-air-ground мереж, питання автоматизації налаштувань наземних базових станцій лишається недостатньо вивченим. Використання методів підкріплювального навчання дозволяє зменшити

## 1. Основна частина

### 1.1. Критичний аналіз літературних джерел

За тематикою статті було проведено пошук літературних джерел за останні 5 років у сфері комп'ютерних наук та інформаційних технологій, які було проаналізовано нижче.

У [1] була запропонована RL модель, що змінює висоту та положення автономного літаючого засобу для надання безпроводного зв'язку. Зміна цих параметрів впливає на якість зв'язку абонентів та точність локалізації. Це не є прямою оптимізацією параметрів мережевих параметрів для статичних базових станцій, а більше навігаційна модель, що націлена на оптимізацію позиціонування літаючих засобів комунікації.

Дослідження [2] пропонує оптимізацію використання загального каналу для формування зв'язку між багатьма базовими станціями (БС) за допомогою DGN-агентів. Кожна БС тренує власний набір агентів, що вибирають оптимальні дії для координованого багатокористувацького планувальника (CMUS) згідно локальних особливостей. Запропонована архітектура значно зменшує надлишковість інформації та обчислювальну складність при комунікації між БС для створення черги користувачів, що зменшує затримку на утворення каналу спілкування між абонентами, однак ніяк не впливає на якість зв'язку між ними.

Джерело [3] представляє спроектований RL агент для мережі повітря-земля із модифікацією для нестационарних даних. Автори використовують модифікований Марківський процес прийняття рішень із динамічною параметризацією, що дозволив ефективно вивчити та використовувати вибір шлюзів та планування каналів користувачів (GSUS) стратегію у симуляції. Дослідники пропонують рішення, що пришвидшує вибір каналів та планування використання каналів для багатьох користувачів, що зменшує затримки, та потенційно має покращити якість зв'язку, однак тільки для мереж типу повітря-земля. У випадку наземних мереж динамічними змінними можуть виступати загальні характеристики сезонності самих даних.

Автори [4] розробили агентів глибокого машинного навчання розподілених по серверах та під'єднанні до контролера мережі, що виконують алгоритм отриманий із загальної моделі, та збирають статистику в режимі близькому до реального часу. Отримані дані використовуються для офлайн навчання загальної моделі, що вивчає загальний оптимальний алгоритм для станцій для розширеного широко-канального мобільного зв'язку та ультра-надійного зв'язку із низькими затримками. Така архітектура вимагає наявності хмарного обчислення для збору, обробки даних та виконання тренування загального агента, а також адаптивного мережевого обладнання із інтелектуальним контролером та відкритим інтерфейсом взаємодії, що може бути дороговартісною модифікацією для провайдерів.

У [5, 6] стверджують, що комунікація для присвоєння каналів з поточним підходом використовує багато надлишкової інформації в процесі і може викликати колізії, що погіршують якість зв'язку та мережі. Ця робота концентрується на оптимізації специфічного типу безпроводного зв'язку - WSN (wireless Sensor Network). Робота покращує Multi-Channel Scheduling Protocol (MCSP) для персональних, безпроводних мереж (Wi-Fi). Фокусується на усуненні колізій при присвоєнні каналів через багатоканальний планувальник. Також присутній другий протокол, це енерго-ефективний RL агент для багатоканальних MAC для MSNs планувальників, що зменшує кількість колізій і розбалансовує залишки енергії на залишкові вузли використовуючи одного RL агента. Третя частина пропонує евристично пришвидшений RL протокол, для присвоєння каналів, що зменшує кількість навчальних ітерацій енерго-ефективним способом враховуючи ширину каналів в планувальнику. Четверта частина роботи пропонує нового кооперативного мульти-агента RL, використовуючи пришвидшену модель для подолання надлишкової комунікації кооперативного агента RL, використовуючи методи само-планування і енергетичного планування. Пропонується дві варіації алгоритму для статичного та динамічного варіанту виконання. Автор заявляє про

покращення результатів пропонованими підходами у різних випадках при декількох експериментах. Однак всі підходи покращують лише для встановлення зв'язку, але ніяк не покращує сам зв'язок. До того ж, розрахована на значно вищі частоти, що за стандартом використовується переважно для інтернет комунікації.

## **1.2. Постановка проблеми та її обґрунтування**

Згідно проведеного аналізу наукових джерел було виявлено ряд проблем:

1. Безпосередню якість голосового зв'язку мереж не приділена достатня увага у роботах.

2. Необхідна оптимізації наземних GSM мереж.

Більшість робіт присвячених оптимізації ресурсів не розглядають як одну з можливих цілей, це якість зв'язку між абонентами. Саме тому, цьому питанню буде приділена увага у цьому дослідженні.

## **1.3. Формулювання мети і задач дослідження**

На основі проведеного критичного аналізу наукових джерел було виявлено, що існує необхідність у оптимізації параметрів GSM мереж, та зокрема якості зв'язку через оптимізацію ресурсів.

Мета роботи полягає в покращенні якості зв'язку динамічним вибором параметрів мережі.

Для досягнення мети необхідно вирішити ряд задач, а саме:

1. Побудова моделі прогнозування стану мережі від зміни параметрів базової станції.
2. Побудова середовища для тренування агента
3. Оцінка якості зв'язку із динамічним підбором параметрів.

**Об'єкт дослідження** – зміна якості зв'язку при динамічному підборі параметрів мережі

**Предмет дослідження** – методи МН оптимізації параметрів GSM мережі із застосуванням.

## **1.4. Розробка оточення та архітектури**

Дані які описують стан базової станції на один день містять наступні колонки: 'Half-Rate Usage Rate', 'Channel Blocking Rate', 'Number of Available Channels', 'Channel Traffic', 'Param\_1', 'Param\_2'. 'Half-Rate Usage Rate' та 'Channel Blocking Rate' це використання половини діапазону каналу для зв'язку та відсоток заблокованих каналів через зайнятість відповідно. Ці значення будуть змінюватися в залежності від виставлених параметрів мережі, тому потребують моделі прогнозування. Кількість вільних каналів та значення трафіку це історичні дані та залишаються не змінними.

На якість голосового зв'язку впливають 2 параметра базової станції ('Param\_1', 'Param\_2'), які визначають порогові значення за яких вмикається, чи вимикається режим половинного каналу зв'язку. Як зрозуміло з назви, на одного абонента припадає половина каналу зв'язку, а отже гіршою якості звук. Другий параметр відповідає за поріг навантаження, при якому повертається використання повного спектру каналу на одного абонента.

Таким чином ми бачимо, що при зміні параметрів мережі, стан цієї ж мережі змінюється, а отже необхідно необхідно це враховувати. Зважаючи на це, було обрано

навчання із підкріпленням, що дає можливість ітеративно оновлювати середовище від взаємодії з ним агента.

Для передбачення майбутніх станів середовища необхідна модель, яка здатна достатньо точно прогнозувати значення на довгий горизонтальний період. N-Hits представлена у [7] дозволяє на основі коваріатів прогнозувати змінні параметри мережі. Таким чином ця мережа відповідає за оновлення середовища агента кожен раз, як його дії відрізняються від історичних.

Оскільки агент має справу із середовищем у 6 змінних та має лише два параметра взаємодії, а також неперервний простір дій (оскільки значення обох параметрів у межах від 0 до 1.0), то було обрано архітектуру м'якого критик представлена у [8].

### 1.5. Результати

Модель прогнозування стану N-Hits була натренована із RMSE похибкою 0,052. На основі цієї моделі вже було запущено тренування агента.

Агент був отриманий шляхом тренування протягом 100 епох по 4 тисячі ітерацій. Після цього його було застосовано до тестувальних даних, з інших БС. В результаті було отримано трохи вищі значення блокування абонента в межах похибки, як зображено на гістограмі на Рисунок 1.

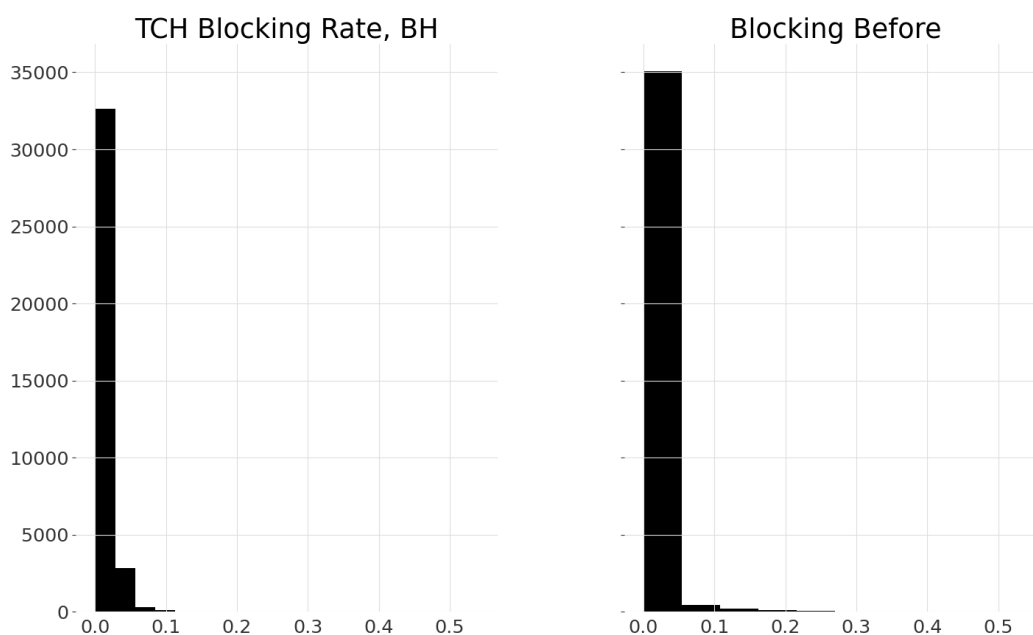


Рисунок 1. Гістограма кількості блокування при застосуванні агента (зліва) та перед ним (справа).

Як зображено на Рисунок 2, якість суттєво покращилася, як видно із розподілу значень використання половинного каналу. Значення змістили свій розподіл, де мода при налаштування агента між 60% та 80% каналів у порівнянні із 90% та 100%. Також ручні налаштування призводили до використання всіх (100% HR) каналів HR-режимі у майже 18000 абонентів, тоді як автоматичні налаштування менше ніж 1000 абонентів.

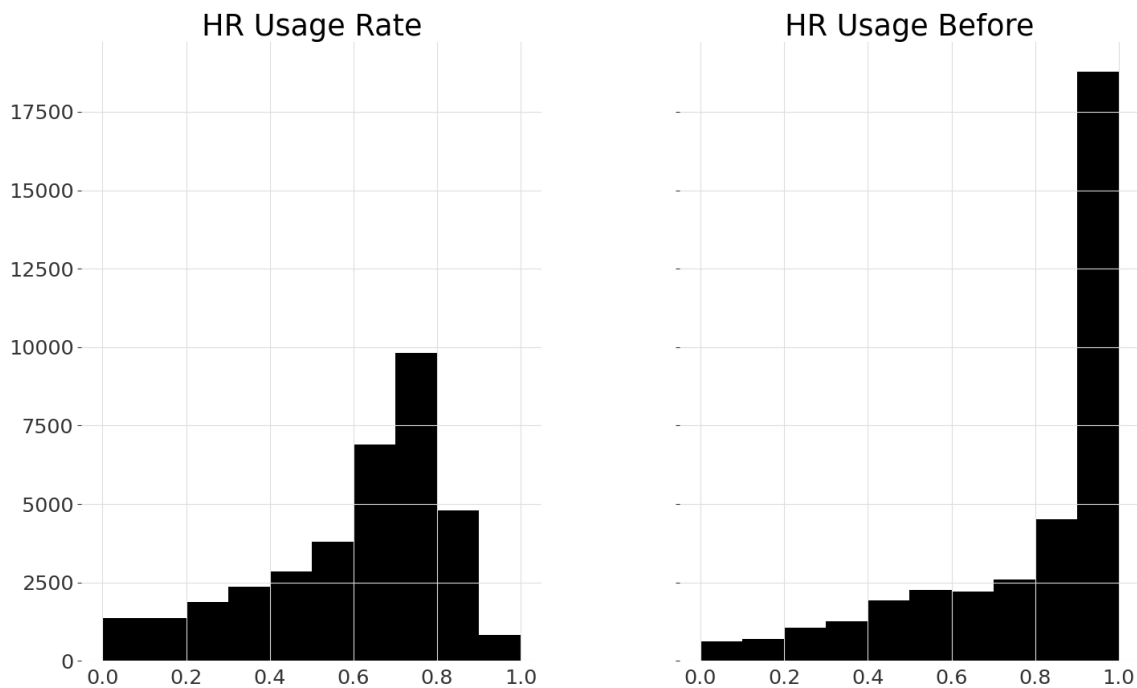


Рисунок 2. Гістограма кількості використання половинного каналу зв'язку при параметрах агента (зліва) та до цього (справа).

Згідно отриманих результатів можна стверджувати, що якість зв'язку для абонентів зросла, оскільки частота та відсоток використання для з'єднання половини діапазону зменшилася. Однак є незначне зростання у кількості блокування з'єднань.

Загалом, автоматичний підбір параметрів на основі RL-агента дозволить зменшити навантаження на польових інженерів операторів мобільного зв'язку, збільшити стабільність та незалежність базових станцій та покращити якість зв'язку між абонентами.

## 2. Висновки

Запропонований підхід до оптимізації ресурсів базових станцій GSM мережі дозволив зменшити використання половинного діапазону каналу на 19.27%, із збільшенням блокування на 0,9%. Автоматичне налаштування параметрів мережі за допомогою RL агента дозволило суттєво покращити якість зв'язку для абонентів незначним збільшенням блокування.

Для подальших досліджень планується зменшити кількість блокування не зменшуючи, чи навіть покращуючи якість зв'язку, а також розробити більш точну модель передбачень наступних станів середовища, для більш надійного агента.

## Список посилань

1. Haile G., Lim J. UAV-BS Operation Plan Using Reinforcement Learning for Unified Communication and Positioning in GPS-Denied Environment. *IEICE Transactions on Communications*. 2024. Bun. E107.B, № 10. С. 681–690.
2. Bai S., Gao Z., Liao X., Sun X. Distributed Adaptive Multiuser Scheduling via Multi-Agent Reinforcement Learning in Multicell MIMO Cellular Networks. 2024.
3. Peng Y., Feng G., Qin S., Wei F., Zhang L. Online Learning Based Joint Gateway Selection and User Scheduling in Non-Stationary Air-Ground Networks. *IEEE Transactions on Cognitive Communications and Networking*. 2024.

4. Alsenwi M., Lagunas E., Chatzinotas S. *Distributed Learning Framework for eMBB-URLLC Multiplexing in Open Radio Access Networks*. *IEEE Transactions on Network and Service Management*. 2024. Bun. 21, № 5. C. 5718–5732.
5. Sahraoui M. *Multi-channel Communication in Wireless Networks*. 2022.
6. KHIDER M. *PHD OF SCIENCE*. University of Biskra, 2022.
7. Challu C., Olivares K. G., Oreshkin B. N., Garza F., Mergenthaler-Canseco M., Dubrawski A. *N-HITS: Neural Hierarchical Interpolation for Time Series Forecasting*. *arXiv*, 2022.
8. Haarnoja T., Zhou A., Abbeel P., Levine S. *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor*. *arXiv*, 2018.

**Рамазанов С. К.** (Східноукраїнський національний університет імені Володимира Даля, м.Київ, Україна)

## КОГНІТИВНІ ТЕХНОЛОГІЇ ПРИЙНЯТТЯ ІНТЕЛЕКТУАЛЬНИХ РІШЕНЬ У НЕЙРОЕКОНОМІЦІ У СКЛАДНОМУ ІНФОРМАЦІЙНОМУ СЕРЕДОВИЩІ

**Анотація.** В умовах цифровізації сучасної економіки та широкого використання систем і технологій штучного інтелекту здатність не просто до творчого мислення, а здатність до неформальних вчинків, неординарних рішень та дій, нововведень – ось головна вимога до фахівців. І тут перше місце виступають теоретичні знання, сприяють розвитку розумового апарату. Навчити розмірковувати самостійно з урахуванням аналізу різних навчань, теорій і поглядів учених-теоретиків стає найпершим завданням щодо і викладання економічної науки. Поведінкова економіка – відносно нова наука та навчальна дисципліна. Вона має швидше міждисциплінарний характер і знаходиться на стику двох наук: психології та економіки, що дозволяє виявити психологічну основу методології прийняття рішень людиною та іншими агентами. Зараз не можна не брати до уваги поведінкову економіку, оскільки поведінковий підхід: притаманний діяльності різних суб'єктів: від індивіда та фірми до ринків та регіонів; систематично повторюється та ускладнюється; дозволяє виявити глибинні мотиви діяльності агентів, які завжди наступних канонам традиційної економіки. Поведінкову економіку донедавна багато хто вважав наукою другорядною, необов'язковою – такою собі неповнолітньою родичкою дорослої традиційної економіки. Однак прихильники останньої все ж таки визнавали, що індивіди час від часу поведуться нерозумно, нераціонально, але на це завжди були теоретичні докази.

**Ключові слова:** поведінкова економіка, нейроекономіка, нейромаркетинг, інтегральна модель динаміки поведінки, поведінкове моделювання СУ та ПР, модель процесу вибору та прийняття рішень, міждисциплінарне дослідження і синтез, економіка інновацій.

**Abstract.** In the context of the digitalization of the modern economy and the extensive use of artificial intelligence systems and technologies, the primary requirement for specialists is not merely creative thinking but the ability to engage in unconventional actions, make extraordinary decisions, and introduce innovations. Theoretical knowledge, which fosters the development of intellectual capacity, takes center stage. Teaching individuals to think independently while analyzing various doctrines, theories, and perspectives of academic scholars becomes a key objective in teaching economic science. Behavioral economics is a relatively new science and academic discipline with an interdisciplinary nature, existing at the intersection of psychology and economics. This allows for uncovering the psychological foundation of decision-making methodologies employed by individuals and other agents. Today, behavioral economics cannot be ignored, as its approach: is intrinsic to the activities of various entities, from individuals and firms to markets and regions; systematically recurs and becomes more complex; and reveals the underlying motives of agents' actions, which often deviate from the traditional economic norms. Behavioral economics was, until recently, considered by many as a secondary, non-essential science—a mere adolescent relative of mature traditional economics. However, even proponents of traditional economics have acknowledged that individuals occasionally act irrationally or unreasonably, and such behavior has always had theoretical explanations.

**Keywords:** behavioral economics, neuroeconomics, neuromarketing, integral behavior dynamics model, behavioral modeling of SU and PR, choice and decision-making process model, interdisciplinary research and synthesis, innovation economy.

**Вступ.** На сьогоднішній день економічний розвиток зазнає певних трансформацій, джерела яких відстежуються у поведінкових реакціях та очікуваннях людського мозку. Саме на основі нового знання про те, як працює мозок, починають змінюватися уявлення про те, як людина приймає рішення. Ці висновки неминуче змінюють уявлення про закономірності та принципи функціонування економіки. Все це зводиться до формування нових зв'язків на стику соціального сприйняття та його економічного результату, створюючи галузь нейроекономіки.

Використання класичних економічних методів в інших науках на сьогоднішній день викликало зворотний ефект – економіка вже приймає безліч методів і концепцій, розроблених психологією, соціологією і навіть ботанікою з механікою. Одним із таких напрямків і є нейроекономіка — галузь науки, яка займається вивченням нейробіологічних засад прийняття рішень.

Нейроекономіка являє собою міждисциплінарний напрямок, який є злиттям економіки, психології та нейронауки у вивченні процесу прийняття рішень людиною. Ця область досліджує основи економічної поведінки шляхом заохочення співробітництва та дискусій між науковцями у галузі психологічних, економічних, неврологічних та інформаційних наук, в тому наук штучного інтелекту.

Нейроекономіка також і нейромаркетинг вивчає процес прийняття рішень при виборі альтернативних варіантів рішень, розподілі ризику та винагороди. Новаторським є те, що саме нейроекономіка використовує економічні моделі для вивчення реакцій людського мозку, а також досягнення нейробиології для побудови економічних закономірностей і моделей. Нейроекономіка є інтердисциплінарним полем досліджень, яке поєднує поняття з економіки, психології та нейробиології. Її мета полягає у вивченні мозкових механізмів, які лежать в основі прийняття рішень, особливо в проблемах маркетингу і складних економічних ситуаціях. Використання інструментів нейроекономіки у прийнятті економічних рішень відкриває нові можливості для глибшого розуміння емоційних та когнітивних факторів, що впливають на поведінку інвесторів, трейдерів та інших агентів ринку. Методи нейроіміджингу та експериментальні ігри дозволяють дослідникам вивчати активацію певних областей мозку у відповідь на різні стимули та ситуації. Аналіз мозкових реакцій на різні активи та ринкові умови дозволяє розробити моделі, що враховують індивідуальні переваги та ризик-профіль маркетологів та інвесторів. Емоції відіграють важливу роль у прийнятті рішень. Дослідження показують, що емоційні реакції можуть проводити оцінку ризику, прийняття рішень про купівлю чи продажу активів, і навіть формування цінових уявлень. Розуміння цих емоційних факторів дозволяє розробити стратегії, які враховують емоційні реакції ринкових агентів.

Розробка інтегрованих ієрархічних інтелектуальних моделей для систем управління і прийняття рішень на основі використання математичних методів, моделей та інноваційних технологій та систем в сучасних умовах нестабільності та кризових явищ є актуальною проблемою. Синтез інтегрованих моделей з урахуванням гуманітарних і когнітивних змінних для оцінки сталого і безпечного розвитку також є важливим і актуальним.

В даній статті вперше розглянуто узагальнена 3-х рівнева структура ієрархічної інтегральної інтелектуальної системи управління і прийняття рішень у складних системах. При цьому враховується: вплив зовнішнього середовища на об'єкта управління (ОУ) з двома виходами («корисним» і «шкідливим»), ОУ як зовнішнє середовище для особи прийняття рішень (ОПР), запропоновано інтелектуальна супервізорна система, система обробки, моделювання, управління та прийняття рішень, з урахуванням можливості моделювання всіх рівнів системи, моделі складного каналу вимірювань та спостережень у вигляді мультиплікативно – адитивної суміші випадкових процесів та факторів, з урахуванням вектора з набору: моделі складного каналу вимірювань та спостережень у вигляді мультиплікативно – адитивної суміші випадкових процесів та факторів, з урахуванням вектора з набору: *когнітивні, емоційні, рефлексивні, духовні та етичні змінні* («параметри порядку»), а також моделі динаміки представлені і в розподілених координатах.

Узагальнена структура ієрархічної інтелектуальної інтегральної системи управління та прийняття рішень у складних системах (рис. 1).

*Позначення на рис. 1:*  $V/C$  – Вимірювання/Спостереження;  $G(t)$  – ціль (цільова установка) Супервізора;  $U(t)$  - прийняте рішення;  $u(t)$  – вектор управління;  $w(t)$  – вектор зовнішніх збурень;  $R(t)$  - ресурси СУ та ПР;  $z(t)$  - вектор "шкідливого" виходу;  $\xi(t)$  – вектор перешкод/випадкових зовнішніх обурень Вимірювання/Спостереження вектор «корисного» виходу;  $y(t)$  – вектор «корисного» виходу;  $Y(t)$  – вектор вимірювання та спостереження вектора стану системи прийняття рішень (СПР);  $X(t) = [k(t), e(t), r(t), \delta(t), \epsilon(t)]$  – вектор стану



системи динаміки поведінки (інтелектуальної СПР), причому  $k(t)$ ,  $e(t)$ ,  $r(t)$ ,  $\delta(t)$ ,  $\epsilon(t)$  – когнітивні, емоційні, рефлексивні, духовні та етичні змінні («параметри порядку»). Причому, вектор  $X(t, q) = \{x_i(t, q), i=1, \dots, n\}$  – це просторові змінні стани інтелектуальної СПР або Суб'єкта управління;  $x(t)$  – вектор стану ОУ.

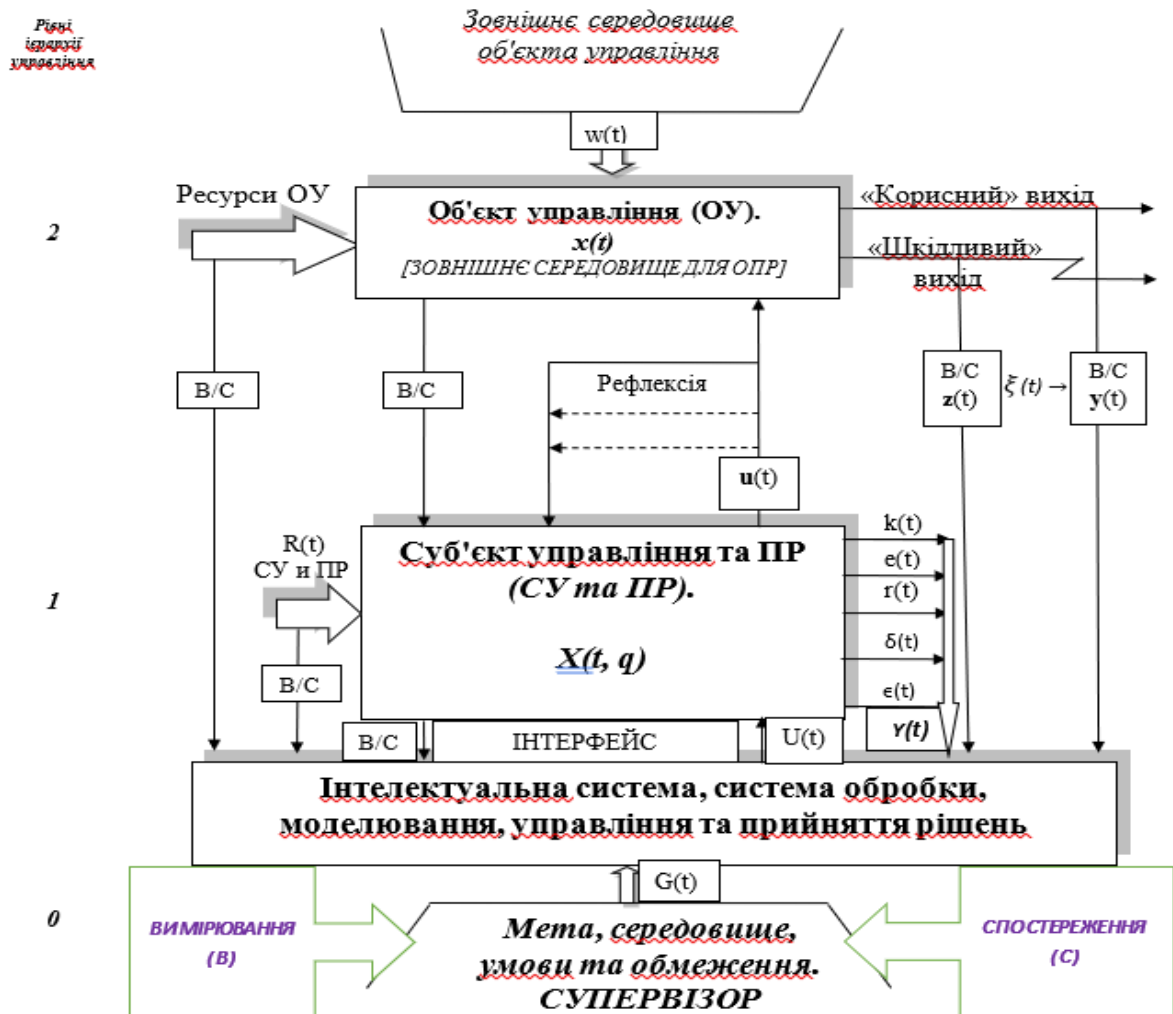


Рисунок 1. Узагальнена структура ієрархічної інтелектуальної інтегральної системи управління та прийняття рішень

### 1. Місце Когнітології в технологіях та системах штучного інтелекту для прийняття управлінських рішень.

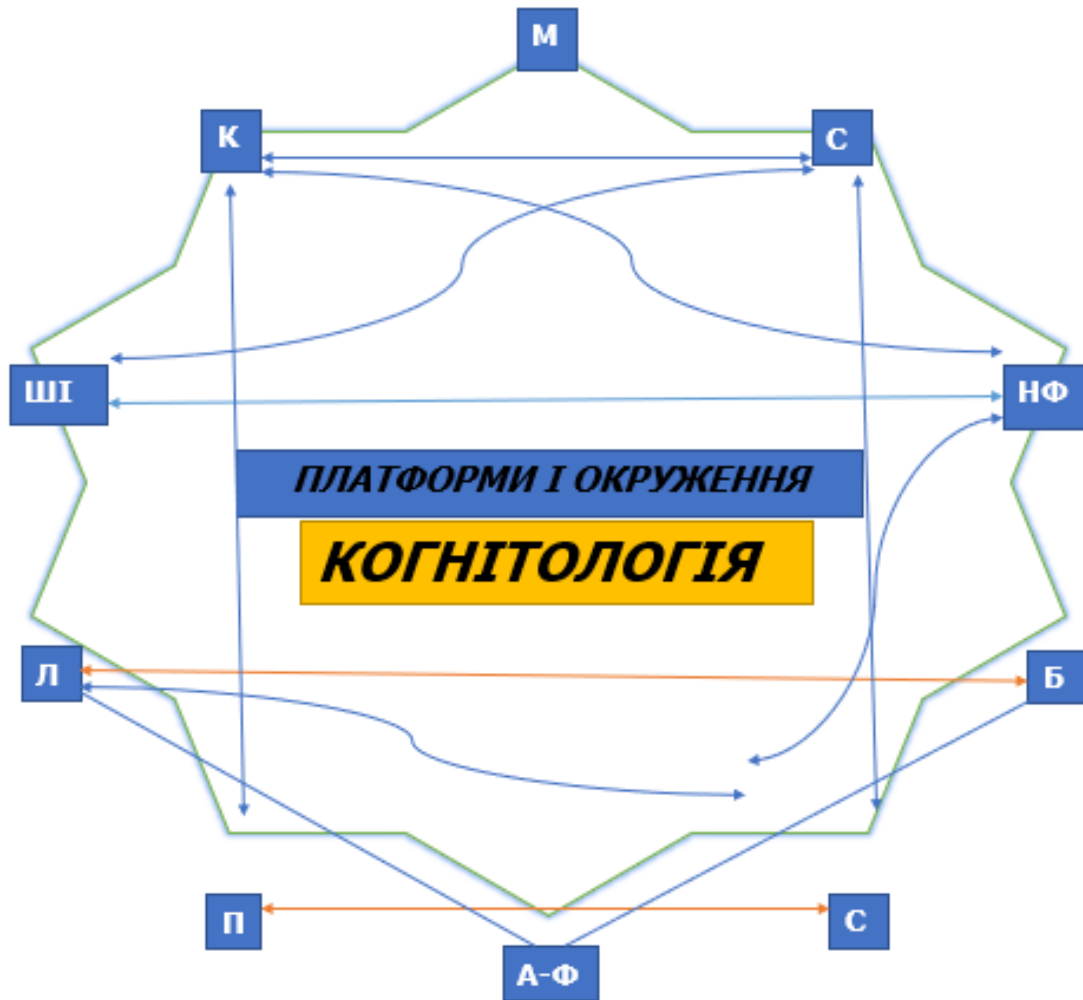


Рисунок 2. Узагальнена схема: Когнітологія та штучний інтелект.

Ця узагальнена схема є кортеж:  $\langle \Phi, \Pi, C, L, A, \text{НФ}, \Pi, B, M, K, C \rangle$ , де:  $\Phi$  - Філософія,  $\Pi$  - Психологія,  $C$  - Соціологія,  $L$  - Лінгвістика,  $A$  - Антропологія,  $A-\Phi$  – Антрополого-філософська підсистема (компонента),  $\text{НФ}$  – Нейрофізіологія,  $\text{ШІ}$  – Штучний інтелект (artificial intelligence,  $\text{AI}^1$ ),  $B$  – Біологія та Біологічні дослідження,  $M-K-C$  – Математика-Кібернетика-Синергетика, основні інструменти наукового пізнання та розвитку.

<sup>1</sup>What is Artificial Intelligence? Архивная копия от 18 ноября 2015 на Wayback Machine FAQ от Джона Маккарти, 2007

## 2. Модель стохастичної динаміки ОУ у просторі станів

Помітимо також, що динамічна система має синергетичний (або кібернетичний) опис, якщо ефективно побудований оператор  $D$ , є такий, що стан системи в кожен момент часу  $t \in (t_0, T(t_0))$  може бути побудований по значеннях вектору  $x(\tau), \tau \in (t_1, t_0)$  за умови, що усі зовнішні дії, що управляють, фіксовані [10, 11]:

$$x(t) = D(x(\tau), \varepsilon, \eta, u), t \in (t_0, T(t_0)), \quad \tau \in (t_1, t_0), \quad (1)$$

де  $\varepsilon(t, r)$ - випадкова дія з відомими імовірнісними характеристиками  $\eta(t, r) \in G_\eta$  дія, задана мірою невизначеності  $G_\eta, u \in R^k$  - дії, що управляють,  $r$  - просторова змінна.

Загальний вигляд моделі динаміки ОУ:  $dx(t)/dt=F(x(t), u(t), w(t), a(t), t)$ . (У разі техногенного промислового ОУ, тобто рівняння динаміки при обліку також екологічних змінних/факторів, див. у роботах автора [1-6]).

Під фазовим станом системи будемо розуміти вектор  $x = (x_1, x_2, \dots, x_n)$ , де  $x_i, i = 1, \dots, n$  – сукупність даних, характеризуючих  $i$ -й сектор системи. Припустимо, що всі сектори оснащені деякими важелями управління:  $u = (u_1, u_2, \dots, u_n)$

Кожен з перерахованих параметрів, в свою чергу, являє собою вектор, складений з числових значень показників різних характеристик динаміки системи. Нехай  $[t_0, t_k]$  – розглянутий період розвитку системи. Позначимо через  $\xi$  вектор деяких додаткових параметрів, які визначаються шляхом прогнозування із застосуванням статистичних даних системи, або вони - константи.

Таким чином, досліджується залежність кожної змінної  $x_i, i = 1, \dots, n$  від управляючих змінних  $u_i$  станів системи і вектора параметрів  $\xi$ .

Зокрема, лінеаризована модель системи має вигляд:

$$\frac{dx(\xi, t)}{dt} = A(\xi)x(\xi, t) + B(\xi)u(\xi, t) + C(\xi)w(t),$$

де  $x(\xi, t) \in R^n, u(\xi, t) \in R^m, w(t) \in R^l, A(\xi) \in R^{n \times n}, B(\xi) \in R^{n \times m}, C(\xi) \in R^{n \times l}, A(\xi) = [A_{ij}(\xi)], B(\xi) = [B_{ij}(\xi)]$  - матриці з параметричною невизначеністю;  $x(\xi, t) = [x_i(\xi, t)]$ - вектор стану системи;  $u(\xi, t) = [u_i(\xi, t)]$  - вектор управління;  $w(t)$  – вектор екзогенних возмущень,  $\xi$  - вектор стохастичних процесів.

Задача синтезу оптимального управління стохастичною динамічною системою ставиться таким чином: знайти закон оптимального управління у вигляді [7-9]:

$$\hat{u}(\xi, t) = N[\hat{x}(\xi, t)] \quad (2)$$

або для лінійного випадку:

$$\hat{u}(\xi, t) = -L\hat{x}(\xi, t), \quad (2^*)$$

а цільовий функціонал для оптимізації (мінімізації) задається у вигляді:

$$J = \int_{\Omega} \left\{ \int_0^{t_k} (x^T(\xi, t)\Theta x(\xi, t) + u^T(\xi, t)Ru(\xi, t))dt \right\} p(\xi)d\xi, \quad (3)$$

де  $N$  - нелінійний оператор.

Тобто після усереднювання маємо критерій оптимального управління як

$$\bar{J}(u) = E \left[ \int_0^{t_k} (x^T(\xi, t)\Theta x(\xi, t) + u^T(\xi, t)Ru(\xi, t))dt \right] \Rightarrow \min. \quad (3^*)$$

У простому випадку можна використати лінійне рівняння спостереження у вигляді:  $y(t) = H(t)x(t) + \eta(t)$ .

У цьому рівнянні спостереження  $H(t)$  є стохастична матриця спостереження [7], яку можна задавати з практичних міркувань, а  $\eta(t)$  є процес типу «білий шум», який може задаватися за даними статистики. При цьому за принципом розділення, треба вирішити два самостійні завдання [8-10].

У ранніх роботах автора розділу завдання оптимального оцінювання (фільтрації) і ідентифікації в мультиплікативно-адитивних сумішах вирішена використовуючи інтегральний опис фільтру [7]. Проте, завдання фільтрації можна і зручно вирішувати також, використовуючи фільтр Калмана для лінійного випадку і фільтру Стратоновича для

нелінійної правої частини рівняння динаміки [7-10]. Оцінки стану знаходяться як умовні середні з побудовою і рішенням нелінійних дисперсійних рівнянь типу Рікати. Фільтр дає рішення стохастичного диференціального рівняння. Він може бути реалізований у вигляді відомої системи із зворотним зв'язком по щільності розподілу початкових умов. Отримана оцінка стану  $\hat{x}(t)$  використовується при рішенні другої задачі - оптимального управління типу  $\hat{u}(t) = -L(t)\hat{x}(t)$ . Як вже відзначалося, завдання оптимального управління зазвичай ставиться на основі принципу Беллмана або принципу максимуму [7-10]. У якості критерію можуть застосовуватися різні функціонали якості, а також вектор функціоналів.

### 3. Модель нелінійної динаміки поведінки ОПП.

*Узагальнена модель динаміки стану/поведінки суб'єкта ПР, тобто, ОПП*

*Модель динаміки у просторі часу:*

$$\frac{d}{dt}X_i(t) = \lambda_{X_i}^{-1}X_i(t) \left[ S_{X_i}(R_{X_i}) - \sum_{j=1}^{N_{X_i}} \beta_{ij}(R_{X_i})X_j(t) \right] + X_i(t)\eta_{X_i}(t) + w_{X_i}(t),$$

$$i = 1, \dots, N_{X_i}$$

$$\frac{d}{dt}R_{X_i}(t) = \lambda_{R_{X_i}}^{-1}R_{X_i}(t) \left[ \overline{R_{X_i}} - \left( R_{X_i}(t) + \sum_{j=1}^{N_{X_i}} \varphi_{X_i X_j}(I, S, D)R_{X_j}(t) \right) \right],$$

$$i=1, 2, \dots, N_{X_i}.$$

*Просторова модель динаміки поведінки ОПП:*

$$\frac{d}{dt}X_i(t, q) = \lambda_i \eta_i(t, q) X_i(t, q) \left[ S_{X_i}(R_{X_i}) - \sum_{j=1}^{N_{X_i}} a_{ij}(R_{X_i}) \prod_{k=1}^j X_k(t, q) \right] +$$

$$+ \sum_{l=1}^3 d_{il} \frac{\partial^2 X_i(t, q)}{\partial q_l^2} + w_i(t, q) + b_i u_i(t), \quad X_i(0, q) = x_{i0}(q), \quad i = \overline{1, n},$$

$$\frac{d}{dt}R_{X_i}(t) = \lambda_{R_{X_i}}^{-1}R_{X_i}(t) \left[ \overline{R_{X_i}} - \left( R_{X_i}(t) + \sum_{j=1}^{N_{X_i}} r_{X_i X_j}(I, S, D)R_{X_j}(t) \right) \right],$$

$$i=1, 2, \dots, N_{X_i},$$

де  $X(t, q)$  – вектор стохастичних просторових змінних станів складної системи (наприклад, стан «свідомості», тобто «мозку»);  $q = (q_1, q_2, q_3)$  – точка у просторі станів (наприклад, координати точки у полі «мозку»),  $\langle \eta_i, w_i \rangle$  – стохастичні мультиплікативні та адитивні обурюючі складові моделі;  $\{a_{ij}(R_{X_i})\}$  – параметри взаємодії між підсистемами (модами);  $\{d_{il}\}$  – коефіцієнти дифузії, що визначають рівень розподілу змінних станів;  $S_X(R)$  – сумарне максимальне (гранично допустиме) значення вектору  $X$ ;  $\{\lambda_{X_i}^{-1}\}$  – сукупність параметрів/факторів, які призводять до хаотичності.

Тут виникає відоме зворотне завдання: оцінювання (фільтрації) та ідентифікації, тобто. визначення оцінки  $\hat{X}(t_k, q^*)$  вектора  $X(t_k, q^*)$  на основі поетапної обробки набору спостережень і при фіксованому  $q^* \in Q$ , тобто.  $\{y(t, t_k) | t \in [t_0, t_k], t \in [t_k, t_{k+1}], \dots\}$  для адаптивного управління, а потім і синтезу самого вектора управління  $U_k \equiv U(t_k) = L[\hat{x}(t_k, q^*)]$ ,  $k = 1, 2, \dots$  [7 - 9]

### 4. Модель динаміки каналу вимірювань / спостережень:

а) Поведінка людини/суб'єкта залежить від набору факторів: <Біологічного, Соціального, Психологічного, Когнітивного, Філософського, Правового та ін.>.

Мозок людини може створити «нове», тобто. що не існує раніше. Мозок структурно перебудовується та стохастично. Сигнали через синапси передаються в обидва напрями, що неможливо фізично.

б) Рівні/якість/форми/типи Свідомості: первинна свідомість (70%), вторинна (10%) та третинна (1%). Населення держави розподілене як: 7:2:1. Третинне мають Генії/Святі. Це біологічна свідомість!

с) Функція поведінки (ФП) ЛПР:

$$Y(t) \equiv P(t) = \langle \text{Моди мозку/функціональні змінні порядку, Змінні ЗС} \rangle.$$

«Моди мозку/функціональні змінні порядку» - це змінні:  $k(t)$ ,  $e(t)$ ,  $r(t)$ ,  $\delta(t)$ ,  $\epsilon(t)$  та інші, а «Змінні ЗС» - це контрольовані тобто, що задаються конструктором систем, але в основному, змінні внутрішнього та зовнішнього середовища (ЗС) – це неконтрольовані/ невизначені/ стохастичні/ непередбачувані обурення:  $\zeta(t)$ ,  $\xi(t)$ .

При цьому важливо розгляд та облік, зокрема, у моделі спостережень/вимірювань як мультиплікативне-адитивна суміш процесів (МАС), наприклад,  $y(t) = h[X(t), \zeta(t), \xi(t), t]$  або, наприклад,

$$y(t) = H(\zeta(t), t)X(t) + \xi(t).$$

Це як аналог, наприклад, виробничої функції чи рівняння спостережень/вимірювань стан складної системи.

Визначення каналу вимірювань як інтегральний оператор:

$$y(t) = \int_{t_0}^t K(t, \tau)X(\tau)d\tau + \xi(t),$$

або інтегральне визначення каналу вимірювань:

$$y(t) = \int_{t_0}^t K(t, \tau)[\zeta(\tau) X(\tau)]d\tau + \xi(\tau).$$

*Модель динаміки каналу спостережень/вимірювань–рівняння спостережень / вимірів поведінки ОПР (СУ):*

$$y(t, t_k) = \int_{t_0}^{t_k} W(t, \tau)[\zeta(\tau) X(\tau)]d\tau + \xi(t), \quad k = 1, 2, \dots,$$

Тут запроваджено позначення, тобто. Визначення нового вектору:

$$[\zeta(t)X(t)] = (\zeta_1(t)X_1(t), \dots, \zeta_n(t)X_n(t))^T, \text{ де } \zeta(t) = (\zeta_1(t), \dots, \zeta_n(t)).$$

Виникає відоме зворотне завдання: оцінювання (фільтрації) та ідентифікації, тобто. визначення оцінки  $\hat{X}(t_k)$  вектору  $X(t_k)$  на основі поетапної обробки набору спостережень  $\{y(t, t_k) | t \in [t_0, t_k], t \in [t_k, t_{k+1}], \dots\}$  для адаптивного управління, тобто самого вектору управління  $u_k = L[\hat{X}(t_k)]$ ,  $k = 1, 2, \dots$ , де  $k$  - номер періоду адаптації,  $k = 1, 2, \dots$ ;  $[t_0, t_k]$ ,  $[t_k, t_{k+1}]$ , ... - інтервали часу адаптивного управління; де  $W(t, \tau)$  - імпульсно-перехідна матриця (ядро інтегрального оператора);  $X(t) = (X_1(t), \dots, X_n(t))$  – стохастичний вектор (стовпець) стану СУ;  $\zeta(\tau)$ ,  $\xi(t)$  – відповідно, стохастичні процеси внутрішніх та зовнішніх обурень на поведінку ОПР, тобто у вигляді моделі «МАС» для адаптивного оцінювання [7-10].  $X(t) = [k(t), e(t), r(t), \delta(t), \epsilon(t)]$  – вектор динаміки стану поведінки СУ (інтелектуальної СПР), причому  $k(t)$ ,  $e(t)$ ,  $r(t)$ ,  $\delta(t)$ ,  $\epsilon(t)$  – когнітивні, емоційні, рефлексивні, духовні та етичні змінні («параметри порядку»). Вектор  $X(t, q) = \{X_i(t, q), i=1, \dots, n\}$  – це набір просторові змінні стану інтелектуальної СПР або Суб'єкта управління, тобто, вектор стану мозку.

## 5. Моделювання адаптивної динаміки поведінки СУ та ПР (ОПР)

Для  $t \in [t_0, T]$ , тобто для всього інтервалу часу функціонування і розвитку системи, розподілену стохастичну модель можна представити як систему диференціальних рівнянь:

$$\partial X_i / \partial t = \Lambda_i [\xi_i (r_i X_i - \sum_{j \neq i} b_{ij} X_i X_j - a_i X_i^2) + D_i(x, y) \Delta X_i] + \zeta_i + u_i,$$

де  $X_i$  - координати стохастичного вектору стану системи, причому  $X_i \equiv X_i(t, q_1, q_2, q_3)$ ;  $i, j = 1, 2, \dots, n$ ;  $r_i$  - коефіцієнт репродукції (розмноження, зростання, розвитку і тому подібне);  $a_i$  - параметр насичення, обмежуюче зростання (репродукцію);  $b_{ij}$  - параметр взаємодії між підсистемами;  $D_i(x, y, z)$  - коефіцієнт дифузії  $i$ -ї підсистеми (суб'єкта, агента) в точці  $q = (q_1, q_2, q_3)$ ;  $\xi_i \equiv \xi_i(t, q)$  і  $\zeta_i \equiv \zeta_i(t, q)$  - стохастичні мультиплікативні і адитивні складові моделі, відповідно;  $u_i \equiv u_i(t, q)$  - координати вектору управління  $u(t)$ , тобто вектор управлінських рішень;  $\Lambda_i$  - масштабуючи коефіцієнт  $\Delta$ -лапласіан, тобто  $\Delta(*) = \partial^2(*) / \partial q_1^2 + \partial^2(*) / \partial q_2^2 + \partial^2(*) / \partial q_3^2$  а  $t \in [0, T]$  - інтервал часу функціонування і розвитку системи.

Зазвичай використовувані *нелінійні функції поведінки* (НФП) ОПР мають статичний характер, тобто. вони не відповідають реальній динаміці функціонування мозку, особливо в критичних та складних ситуаціях (КСС) та умовах. Тому актуальними є дослідження НФП, в якому враховуються як динаміку, так і стохастичність впливу ЗС, наприклад, у вигляді [7]:

$$y(t) = \int_{t_0}^t K(t, \tau) x(\tau) d\tau + \xi(t) \quad \text{або} \quad y(t) = \int_{t_0}^t K(t, \tau) \zeta(\tau) x(\tau) d\tau + \xi(t),$$

де  $K(t, \tau)$  - імпульсно-перехідна матриця (ядро інтегрального оператора);  $x(t) = (x_1(t), \dots, x_n(t))$  - вектор (стовпець) ресурсів («вхід» системи);  $\xi(t)$  і  $\zeta(t)$  - стохастичні процеси впливів, що обурюють (екзогенні та ендогенні) на поведінку ОПР.

Поведінка як функція/змінна залежить від багатьох факторів/змінних свідомості та підсвідомості: когнітивні, емоційні, рефлексивні, етики, норм моралі/моралі (совісті) - релігійних заповідей, норм права, рівня культури та освіти та ін. Зауважимо, що деякі з них швидко, а інші - повільно змінюються.

Причому одні мають динаміку (нелінійну), а інші - просто деякі функції часу в обмеженнях. Совість і моральні цінності входять до етичної складової, тобто. треба моделювати етичні змінні, тобто. процеси етики. Вони варіабельні.

Поняття етичний ІІІ. За кордоном немає наших зазначених понять, моральних цінностей та інші духовні фактори. Навіть для деяких з них немає перекладу, з чим іноді страждаю в публікаціях.

*Як мозок приймає рішення.* Багато факторів, що впливають на прийняття рішень (гени, соціум та інші), серед них мозок - ключовий механізм, який пояснює наші рішення. Так виникла нова галузь - нейроеконіміка, яка намагається створити принципово нову теорію, яка пояснить наші рішення генами, активністю нейронів, сприйняттям нашим мозком інформації, впливом соціального середовища та еволюцією.

Ні для кого не секрет, що дослідження мозку - складний процес: мозок складається з мільярдів нейронів, які об'єднані десятками тисяч зв'язків, і зрозуміти, як працює ця конструкція непросто, проте нейробіологія дійшла досить серйозного прогресу у вирішенні цього питання.

Дослідження Бенджаміна Лібета сильно вплинуло на нейробіологів і філософів: дослідник отримав можливість хронометрувати свідомість рішень. Лібет зареєстрував енцефалограму, щоб подивитися, що відбувається в мозку людини в момент прийняття рішення, і виявив, що усвідомлення рішення належить самому дії, що цілком логічно, проте дослідник звернув увагу на те, що сигнал мозку істотно передбачає усвідомлення рішення, і

можна передбачити рішення людину задовго до моменту, коли вона її усвідомлює. За допомогою приладу за півсекунди до усвідомлення рішення воно вже ставало відомим.

Це відкриття породило величезну дискусію у багатьох фахівців різних спрямованостей, особливо у філософів, які запитали про те, наскільки обґрунтовано думку про те, що саме людина приймає рішення. Повторне дослідження з використанням МРТ показало, що за вісім секунд до того, як рішення стає усвідомленим, його можна передбачити.

У нейроекономіці прийняття рішень виглядає так: виникає проблема, слідує порівняння альтернатив, відбувається вибір між альтернативами та оцінюється результат з метою отримання досвіду.

Більшість досліджень у нейроекономіці фокусуються на етапі порівняння альтернатив.

Вважається, що, якщо відома перевага, можна передбачити рішення. Тому більшість досліджень сконцентрована на виявленні, чому нам подобається одна опція більше, ніж інша. Для вирішення цього питання можна скористатися економічною теорією - правилом максимізації корисності, згідно з яким, вибираючи між двома варіантами, ми вибираємо той, який нам більше подобається або, згідно з економічною теорією, має найбільшу корисність (привабливість опції).

Однак що має велику цінність? Для класичної науки визначити цінність майже неможливо – економіці абсолютну цінність ніколи не визначають.

Дослідники мозку кажуть, що проблему можна вирішити. З точки зору нейроекономіки, суб'єктивні цінності – це середнє значення активності нейронів у спеціалізованих галузях мозку, що кодують наші уподобання.

Ієрархічна супервізорна суб'єктно- та об'єктно-орієнтована адаптивна інтелектуальна система управління та ПР основі інтегральної стохастичної моделі нелінійної динаміки можна уявити у вигляді (рис. 3).

*За підсумками інтегральної стохастичної моделі нелінійної динаміки.*



Рисунок 3. Ієрархія управління та ПР: СУПЕРВІЗОР → ОПР ОУ.

**ВИСНОВКИ:** Вперше розглянуто узагальнена 3-х рівнева структура ієрархічної інтегральної інтелектуальної системи управління і прийняття рішень у складних системах, запропоновано інтелектуальна (супервізорна) система, система обробки, моделювання, управління та прийняття рішень, з урахуванням можливості моделювання всіх рівнів системи, моделі складного каналу вимірювань та спостережень у вигляді мультиплікативно – адитивної суміші випадкових процесів та факторів, з урахуванням вектора з набору: моделі складного каналу вимірювань та спостережень у вигляді мультиплікативно – адитивної суміші випадкових процесів та факторів, з урахуванням вектора з набору: *когнітивні,*

емоційні, рефлексивні, духовні та етичні змінні («параметри порядку»), а також моделі динаміки представлені і в розподілених координатах.

### Список посилань.

1. Рамазанов С. К., Бурбело О. А., Вітлінський В. В. и др. Ризики, безпека, кризи і сталий розвиток в економіці: методології, моделі, методи управління та прийняття рішень. Монографія. / Під заг. ред. проф. С.К. Рамазанова. – Луганськ: Вид-во «Ноулідж», 2012. – 948 с.
2. Рамазанов С.К. Об'єктно + суб'єктно орієнтований підхід в управлінні техногенної виробничої системою в умовах невизначеності // Вісник СНУ ім. В. Даля, № 2[156], ч. 1, 2011 – С. 251-258.
3. Економічна кібернетика: моделювання соціально-економічних систем: колективна монографія / за заг. ред. Л.М. Савчука, К.Ф. Ковальчука. – Дніпро: Пороги, 2017. 480 с.
4. Галіцін В.К., Рамазанов С.К. Інтегральна стохастична нелінійна модель динаміки інноваційної економіки// Науково-аналітичний журнал «Моделювання та інформаційні системи в економіці» / Сб. наук. праць / Головн. ред. В.К. Галіцін. – Київ: КНЕУ, 2017. - С. 50-64.
5. Рамазанов С.К. Прогнозування розвитку інноваційної економіки на основі інтегрованої стохастичної моделі динаміки зростання. - С. 146-153// Актуальні проблеми прогнозування поведінки складних соціально-економічних систем: Монографія / За ред. О.І. Черняка, П.В. Захарченка. – Бердянськ: Видавець Ткачук О.В., 2017. – 371с.
6. Sultan Ramazanov, Oleksandr Chernyak, Bogdan Tishkov, Renat Ahmedov and Oleksandr Honcharenko. Non-linear forecasting of the state of a socio-eco-oriented innovative economy in the conditions of systemic crises. – С.164-191. // SHS Web of Conferences Volume 65 (2019). The 8th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2 2019). Odessa, Ukraine, May 22-24, 2019.
7. Рамазанов С.К. Оценивание и идентификация стохастических мультипликативно-аддитивных смесей./Автореф. дисер. на соиск. уч. степ. к.т.н., Киев: ИК АН Украины, 1982.– 24 с.
8. Рамазанов С.К., Рогоза Н.Е., Истомин Л.Ф. Оптимальное управление экономическими системами: учебное пособие / С.К. Рамазанов, Н.Е. Рогоза, Л.Ф. Истомин. – Полтава: ПУЭТ, 2016. – 380с.
9. Рамазанов С.К., Рогоза Н.Е., Истомин Л.Ф. Нелинейная оптимизация и вариационное исчисление (в примерах и задачах): уч. пособ. / С.К. Рамазанов, Н.Е. Рогоза, Л.Ф. Истомин. – Полтава: ПУЭТ, 2016. – 124с.
10. Острем К. Введение в стохастическую теорию управления. – М.: Мир, 1970.– 326 с.



Слюсар В.І. (Центральний науково-дослідний інститут озброєння та військової техніки ЗС України)

## НЕЙРОМЕРЕЖНИЙ АНАЛІЗ СТРУКТУРИ ХМАР НА БОРТУ АВІАЦІЙНИХ ПЛАТФОРМ ДЛЯ ПРОГНОЗУВАННЯ МЕТЕОДАНИХ У ПРИЗЕМНОМУ СЕРЕДОВИЩІ

**Анотація:** Доповідь присвячена розробці системи нейромережного аналізу структури хмар для визначення метеорологічних даних під хмарами на основі зображень, отриманих із борта авіаційних платформ. Актуальність теми полягає у потребі підвищення точності прогнозів погоди у регіонах із обмеженою наземною інфраструктурою. Запропонований підхід передбачає інтеграцію методів комп'ютерного зору та нейронних мереж для вилучення характеристик хмар і визначення метеорологічних параметрів (атмосферний тиск, швидкість і напрямок вітру, тип та інтенсивність опадів). Архітектура системи складається з модуля вилучення ознак із використанням згорткових нейронних мереж (CNN) і регресійного блоку для прогнозування параметрів. Навчання моделі базується на датасеті, що включає зображення хмар різних типів, отримані з бортових камер авіаційних платформ, і відповідні метеодані, зібрані з наземних станцій. Наголошується на важливості забезпечення репрезентативності даних, врахуванні сезонних змін і географічних особливостей. Практичне значення роботи полягає в можливості використання системи для оперативного аналізу погодних умов у режимі реального часу. Це включає прогнозування погоди для авіації, оцінку кліматичних умов для енергетики та сільського господарства. Також розглядаються перспективи застосування великих мовних моделей та мультиагентних систем і узагальнення підходу для інших сфер, таких як моніторинг інфраструктури чи природних катастроф.

**Ключові слова:** нейромережний аналіз, структура хмар, метеорологічні дані, комп'ютерний зір, згорткові нейронні мережі, прогнозування погоди, авіаційні платформи,

**Abstract:** The paper is dedicated to the development of a neural network-based system for analyzing cloud structures to determine meteorological data beneath the clouds based on images captured from aviation platforms. The relevance of the topic lies in the need to improve weather forecast accuracy in regions with limited ground infrastructure. The proposed approach involves integrating computer vision methods and neural networks to extract cloud characteristics and determine meteorological parameters (atmospheric pressure, wind speed and direction, type, and intensity of precipitation). The system architecture consists of a feature extraction module utilizing convolutional neural networks (CNNs) and a regression block for parameter prediction. The model training is based on a dataset that includes images of various cloud types captured by onboard cameras of aviation platforms and corresponding meteorological data collected from ground stations. Emphasis is placed on ensuring the representativeness of the data, accounting for seasonal changes, and considering geographical features. The practical significance of the work lies in the possibility of using the system for real-time weather condition analysis. This includes weather forecasting for aviation, assessing climatic conditions for energy and agriculture. The prospects for applying large language models and multi-agent systems are also considered, as well as generalizing the approach to other areas such as infrastructure monitoring or natural disaster management.

**Keywords:** neural network analysis, cloud structure, meteorological data, computer vision, convolutional neural networks, weather forecasting, aviation platforms.

Сучасні задачі метеорології та кліматології вимагають високої точності прогнозів у реальному часі, особливо в умовах регіонів із обмеженою наземною інфраструктурою для збору даних. Традиційні методи аналізу погодних умов, які базуються на даних із наземних станцій, супутників або радарів, часто не забезпечують достатньої деталізації та оперативності. Водночас, розвиток авіаційних платформ, таких як безпілотні літальні апарати (БПЛА) та літаки, а також повітряних зондів, відкриває нові можливості для збору високоточних даних про хмарність безпосередньо з висоти польоту. Хмари є важливим елементом атмосфери, що прямо впливає на такі метеорологічні параметри, як опади, атмосферний тиск, швидкість і напрямок вітру. Їх структура, форма та динаміка можуть виступати індикаторами погодних змін. Однак аналіз хмар у реальному часі для оцінки приземних метеоданих залишається складним завданням через варіативність хмарності, сезонні зміни та вплив локальних факторів.

Мета роботи полягає у створенні концепції системи, яка здатна на основі зображень хмар, отриманих із борта авіаційних платформ, визначати ключові метеорологічні параметри

в приземному середовищі. Такий підхід вимагає інтеграції методів комп'ютерного зору, глибокого навчання та прогнозування часових рядів.

Дослідження хмарності та її впливу на атмосферні процеси є однією з ключових задач кліматології та метеорології. Вивчення структури хмар дозволяє не лише покращити прогнози погоди, а й сприяє розумінню глобальних кліматичних змін. Використання нейромереж для аналізу структури хмар з метою визначення параметрів приземних шарів атмосфери є новим підходом, що об'єднує дослідження атмосфери та штучного інтелекту, сприяючи розвитку міждисциплінарних напрямів. Практичне значення пропонованого підходу включає оперативний аналіз погодних умов для планування польотів авіації, оцінку регіонального клімату та прогнозування опадів для екології та сільського господарства, планування роботи відновлюваних джерел енергії в енергетиці, а також прогнозування екстремальних погодних явищ для запобігання катастрофам та організації евакуації населення. Розробка системи нейромережного аналізу хмарних покривів із використанням авіаційних платформ дозволить значно розширити доступ до оперативних і точних метеорологічних даних, забезпечуючи якісно новий рівень у вирішенні задач прогнозування та моніторингу погоди.

Аналіз останніх публікацій свідчить про зростаючий інтерес до використання штучних нейронних мереж для аналізу хмарності та прогнозування метеорологічних параметрів. Зокрема, дослідження [1] демонструє ефективність рекурентних нейронних мереж (RNN) та їх комбінацій у прогнозуванні температури навколишнього середовища, досягаючи похибки 14.22%. Це підкреслює потенціал нейронних мереж при вирішенні метеорологічних задач. Інші роботи [2 - 10] присвячені розробці нейромережних моделей для автоматичної класифікації хмарних структур на зображеннях переважно з наземних камер спостереження. Це важливо для розуміння атмосферних процесів та покращення прогнозів погоди. Для цього також використовуються зображення хмар з супутників [11, 12]. У статті [13] розглядається застосування ШІ для аналізу великих обсягів метеорологічних даних, що дозволяє швидко виявляти закономірності та підвищувати точність прогнозів. Ці та інші роботи закладають основу для подальших досліджень у сфері використання нейронних мереж для аналізу та прогнозування метеорологічних параметрів, підкреслюючи важливість інтеграції методів глибокого навчання у метеорологічну практику.

Ідея визначення погодних умов на землі, під хмарами, за допомогою аналізу зображень хмарного покриву, зроблених із літака, виглядає досить перспективною, особливо з урахуванням сучасних можливостей комп'ютерного зору та нейронних мереж [14 - 16]. Зовнішній вигляд суцільного хмарного шару, приклади якого наведено на рис. 1, 2, може надати корисну інформацію про погодні умови, оскільки форма, структура, висота та тип хмар пов'язані з процесами в атмосфері, включаючи опади, тиск і вітер. Це може бути особливо корисним у районах, де немає метеостанцій (рис. 1).



Рисунок 1. Приклад щільного хмарного покриву над Атлантичним океаном (фото автора).

Для реалізації такої ідеї необхідно врахувати кілька важливих етапів. По-перше, потрібно зібрати велику та якісну базу репрезентативних даних. Це мають бути зображення хмар, зроблені з різної висоти, під різними ракурсами і в різноманітних погодних умовах, а також відповідні метеорологічні дані із землі. Вказані дані мають охоплювати усі можливі типи хмарності, пори року, географічні зони та екстремальні погодні явища. Додатково, можна

інтегрувати зазначені дані з іншими джерелами, такими як супутникові зображення, показники атмосферного тиску на різних висотах, метеорадарні та лідарні дані тощо. Це дозволить моделі враховувати не лише вигляд хмар, але й інші важливі фактори, які впливають на погоду. Також доцільно запровадити механізми зворотного зв'язку, коли дані прогнозу, отримані моделлю, зіставляються з реальними спостереженнями, щоб покращити точність прогнозів і адаптивність системи. Наприклад, перспективним напрямом може бути використання подібної системи в реальному часі для авіації. Літак чи БПЛА має оснащуватися сенсорами та камерами, які в реальному часі передаватимуть дані в обчислювальний центр, наприклад, за допомогою стільникового зв'язку, де він є [17], або системи Starlink. Також можливий прийом на борту у вказані способи наземних метеоданих для проведення оперативного аналізу із прогнозуванням погодних умов під літаком. Це дозволить підвищити безпеку польотів, прогнозувати локалізацію зон турбулентності.

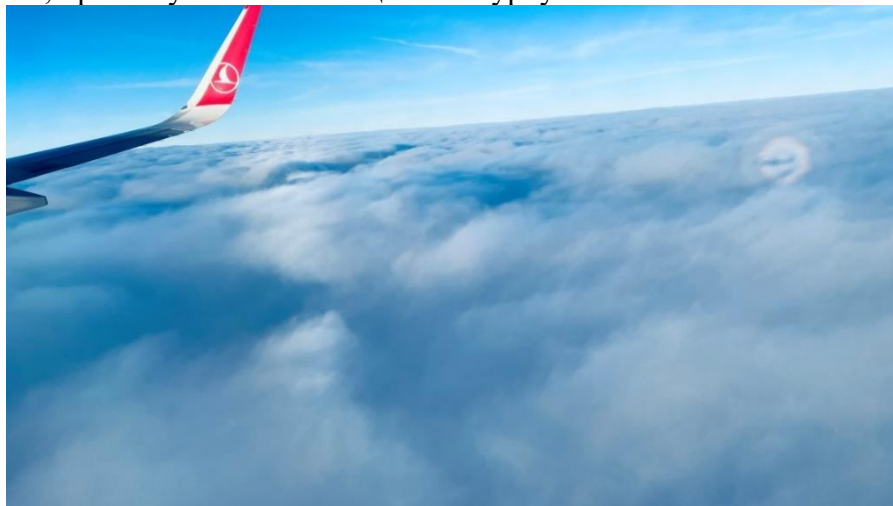


Рисунок 2. Пухкі хмари (фото автора).

По-друге, важливо створити модель нейронної мережі, яка могла б ефективно аналізувати зібрані дані. Для цього доцільно використовувати згорткові нейронні мережі (CNN) для обробки зображень, а також рекурентні мережі (RNN) або трансформери для обробки часових і послідовних даних.

Звісно, прогнозування параметрів погоди на землі під хмарами може мати певну похибку через складність атмосферних процесів. Тому на першому етапі доцільно обмежитися аналізом лише суцільних хмарних покривів, що має спростити встановлення закономірностей еволюції метеоданих. Для цього безпосередньому прогнозуванню атмосферних параметрів через аналіз динаміки візуальних процесів має передувати класифікація стану хмарних шарів. Вирішення даної задачі може бути здійснене аналогічно запропонованому в [18] підходу до класифікації сезонних умов із використанням згорткових нейронних мереж, який можна удосконалити та адаптувати для задач визначення суцільної хмарності. При цьому основою процесу ідентифікації суцільних хмарних станів є формування датасету, що включає зображення різних типів хмарності: суцільної, часткової або її відсутності. Як вже зазначалося, ці зображення отримуються з борта літаків або безпілотних апаратів на різних висотах. Важливо забезпечити різноманітність даних - знімки з різними рівнями освітлення, у різні пори року та з багатьох регіонів, що дозволить зробити модель більш універсальною.

Для класифікації типів хмарності можна використати апробовану в [18] архітектуру на основі MobileNetV2 [19] (рис. 3), попередньо навченої на датасеті ImageNet [20], яка зарекомендувала свою ефективність. Суттєво, що завдання класифікації може бути поєднане з сегментацією хмарного зображення на основі мережі U-Net. Перед аналізом усі зображення проходять попередню обробку, що охоплює зміну розмірів, нормалізацію кольорів і забезпечує уніфікацію даних. Після цього нейронна мережа починає аналізувати хмарне середовище. Вона визначає, чи присутня на зображенні суцільна хмарність, і, якщо так,

виділяє її межі. Крім того, в процесі сегментації можливо уточнити тип хмар - наприклад, перисті, купчасті чи шаруваті. Для досягнення найкращих результатів виконується оптимізація гіперпараметрів, таких як розмір батчу, крок навчання та кількість епох. Наприклад, зменшення розміру батчу показало себе в [18] як ефективний спосіб знизити вимоги до обчислювальних ресурсів, що може бути особливо важливим для умов із обмеженим обладнанням. Після цього модель проходить етап валідації. Як і у випадку класифікації сезонних умов [18], використовуються тренувальні, валідаційні та тестові вибірки. Точність моделі оцінюється за допомогою метрик, таких як Balanced Recall, що дозволяє визначити її ефективність. Завершальним етапом є реалізація системи у реальному часі. Навчена модель інтегрується у бортове програмне забезпечення авіаційних платформ, що дозволяє отримувати результати класифікації та використовувати їх для подальшого передбачення погодних умов.



Рисунок 3. Застосування попередньої класифікації хмарного покриву на вході нейромережі прогнозування приземних метеоданих.

Для безпосереднього визначення метеорологічних даних під суцільними хмарами може бути запропонована архітектура нейронної мережі, що базується на інтеграції сучасних технологій комп'ютерного зору та регресійних моделей. Вона складається з двох ключових компонентів: модуля для вилучення ознак із зображень хмар і модуля для перетворення цих ознак у метеорологічні параметри.

Перший компонент - модуль вилучення ознак, - використовує згорткову нейронну мережу, наприклад, MobileNetV3 або EfficientNetB0, яка оптимізована для роботи в умовах обмежених ресурсів. На вхід системи подаються зображення розміром  $224 \times 224 \times 3$ , які проходять через серію шарів згортки, об'єднання та нормалізації. На виході формується високорівневий набір ознак, що відображає основні характеристики хмарного середовища. Для стиснення цього багатовимірного представлення у компактний вектор використовується додатковий шар об'єднання, а для запобігання перенавчанню застосовується механізм Dropout.

Другий компонент, регресійний блок, складається з одного або кількох повнзв'язних шарів із функцією активації ReLU. Вихід моделі представлений кількома нейронами, кожен з яких відповідає за прогноз окремого параметра: атмосферного тиску, швидкості вітру, напрямку вітру або типу опадів. Для кількісних параметрів використовується функція активації Linear, тоді як для класифікаційних - Softmax.

Вхідні дані для моделі формуються із зображень хмар, які можуть бути попередньо оброблені для покращення контрасту, а також додаткових метаданих, наприклад, геолокації чи висоти польоту. Особливу увагу, знову ж таки, слід приділити сезонним варіаціям. Наприклад, зимові хмари часто здаються яскравішими через відбите від снігу світло, що може вплинути на прогноз. З урахуванням цих змін модель адаптується до різних умов. Вихідні дані представляють собою прогнозовані метеорологічні параметри як у вигляді регресійних значень (тиск, швидкість вітру), так і класифікаційних (тип опадів). Для підвищення стійкості

моделі до різноманітних умов застосовуються техніки аугментації, такі як обертання, зміна яскравості та перевернення зображень.

Навчання моделі базується на використанні функцій втрат Mean Squared Error для регресійних прогнозів і Categorical Cross-Entropy для класифікації. Гіперпараметри моделі включають розмір батчу, початковий крок навчання та оптимізатор. Реалізація архітектури може бути виконана за допомогою популярних фреймворків, таких як TensorFlow/Keras або PyTorch.

Для подальшого розвитку системи передбачено інтеграцію механізмів самонавчання, що дозволить адаптувати модель до нових типів хмарності чи метеорологічних умов. Додатково можливо впровадити модуль Explainable AI, який забезпечить прозорість прогнозів і покращить їх інтерпретованість. Ця архітектура є гнучкою, масштабованою та відкриває нові можливості для точного визначення метеорологічних даних за зображеннями хмар.

У запропонованій архітектурі метеодані при навчанні подаються на вихід, тобто вони виступають цільовими значеннями (targets) для оптимізації моделі, подібно до того, як це робиться в задачах регресії або класифікації. Вони не є вхідними даними, як це могло б бути в задачах, де модель використовує допоміжну інформацію для аналізу. На вхід подаються тільки зображення хмар. Це основні дані, з яких модель буде вчитися вилучати ознаки. На виході модель прогнозує набір метеорологічних параметрів, а саме: регресійні значення, наприклад, тиск, швидкість вітру; класифікаційні категорії (тип опадів, тип хмар, напрям вітру). Ці параметри порівнюються з еталонними значеннями (ground truth) за допомогою відповідних функцій втрат (loss functions).

Суттєво, що для кожного зображення хмар в навчальному наборі даних має бути відповідний набір цільових метеорологічних параметрів, зібраних із наземних метеостанцій або інших джерел (наприклад, супутників). Під час навчання мережа намагається мінімізувати різницю між прогнозованими і реальними значеннями метеорологічних параметрів. Такий підхід дозволяє навчити модель генерувати коректні прогнози метеоданих виключно на основі аналізу зображень хмар, без додаткових входів або потреби в сегментаційній масці.

В розглянутій задачі метеодані є компактними значеннями (набором чисел або класів) і не мають просторової прив'язки, як у сегментаційних задачах. Тому структура моделі скоріше нагадує архітектуру для багатозначного прогнозування (multi-output regression/classification), а не U-Net з сегментації зображень.

Разом з тим, в даному випадку відповідну нейромодель можна розглядати як різновид моделі Image-to-Text, але з адаптацією під специфіку задачі прогнозування числових та категоріальних метеорологічних даних. Як відомо, у задачах Image-to-Text модель аналізує зображення та генерує текстовий опис. У випадку, що розглядається, модель аналізує зображення хмар і генерує структуровані числові (регресійні) та текстові (класифікаційні) дані, що відповідають метеорологічним параметрам. Вихід моделі є узагальненим "текстом", тобто описом погодних умов у вигляді структурованого набору даних.

Як і в класичній Image-to-Text архітектурі, в описаній метеозадачі можливо використовувати сукупність енкодера та декодера. Енкодер (наприклад, MobileNetV3, EfficientNet) перетворює зображення у компактне представлення (feature vector), а декодер працює над цим представленням для генерації вихідних значень. У даному випадку це можуть бути окремі лінійні або класифікаційні шари для кожного параметра.

Вихідні метеодані можуть бути представлені в текстовій формі, наприклад, у форматі yaml, як показана на рис. 2. Генерація таких структурованих даних - це один з можливих результатів роботи моделі Image-to-Text, але тут виходи є чітко визначеними. Для більш універсального підходу можна задіяти Image-to-Text моделі типу CLIP або BLIP, які працюють на стику зображень і тексту. Вони можуть бути перенавчені на задачі "зображення хмар → опис метеоданих". Наприклад, модель генеруватиме текстовий опис: "Суцільні хмари, опади відсутні, швидкість вітру 10-15 км/год, напрямок північно-західний." Переваги такого Image-to-Text підходу полягають у тому, що модель стає більш універсальною і може бути легко адаптована для різних завдань.

Генерований вихід може використовуватися не лише як числові дані, а й як інтерпретований текстовий опис для оператора або системи. Для реалізації зазначених моделей використовуються сучасні фреймворки, такі як PyTorch, з адаптованими компонентами encoder-decoder. Можливе підключення великих мовних моделей (LLM) [21] до кінцевого етапу для генерації описів на природній мові. Цей підхід робить ідею ще більш масштабованою, дозволяючи інтеграцію в багатомодальні системи, де зображення і текст працюють як доповнюючі джерела інформації.

Зокрема раціональним вибором для описаної задачі є використання мультимодальних LLM [21], таких як LLaVa (Large Language and Vision Assistant) або Pixtral та ін. Ці моделі створені для багатомодальних задач і дозволяють безпосередньо працювати із зображеннями, генеруючи текстові або структуровані виходи на основі зорової інформації. Моделі LLaVa і Pixtral інтегрують аналіз зображень та генерацію тексту. Це дозволяє працювати з даними хмар як із вхідною інформацією і отримувати готовий текстовий або числовий опис метеоданих. При цьому немає потреби розробляти спеціальну регресійну модель, оскільки вказані моделі адаптовані до складних вхідних даних та генерації структурованих виходів. Pixtral також може інтегрувати генерацію категорійних (опади, тип хмар) і числових даних у текстовому форматі. З метою адаптації для конкретної задачі LLaVa має бути перенавчена (fine-tuned) на специфічних даних - зображеннях хмар із прив'язкою до метеоданих. Це дозволить адаптувати модель до унікальної задачі, для чого необхідно створити спеціалізований датасет з парами "зображення хмар → метеодані". Такий датасет, як вже зазначалося, має включати зображення хмарних шарів (супутникові, з БПЛА, з літака), а також відповідні дані про атмосферний тиск, вітер, опади, отримані з наземних метеостанцій чи супутникових сенсорів. Зображення подається на вхід моделі (LLaVa або Pixtral), а модель далі генерує текст із описом метеоданих. Вихідний результат може бути наданий як у вигляді готових текстових описів (наприклад, "швидкість вітру 10 м/с, опади відсутні"), так і у форматі JSON для подальшої обробки. LLaVa підтримує пояснення своїх прогнозів і може, наприклад, їх коментувати: "Зображення показує суцільні шаруваті хмари, які зазвичай пов'язані з атмосферним тиском 1005-1010 hPa і вітром у діапазоні 5-10 км/год." Тонке налаштування (fine-tuning) LLaVa дозволить перенавчити її на своєму наборі даних, щоб покращити точність прогнозів. Для Pixtral, враховуючи її значні розміри, замість fine-tuning можна використовувати техніку додавання спеціалізованих шарів для регресійних задач. В свою чергу, Pixtral може використовувати своє розширене розуміння зорової інформації для більш точного аналізу. Після донавчання слід перевірити ефективність LLM-моделі на нових зображеннях, оцінивши її точність, використовуючи метрики MSE для регресійних прогнозів і точність (accuracy) для класифікаційних. Після завершення процесу донавчання модель може бути розгорнута на сервері чи хмарній платформі для аналізу зображень у реальному часі.

Вказані LLM дозволяють автоматично генерувати текстові результати різними мовами, наприклад, українською чи англійською, що зручно для інтеграції в глобальні системи. З урахуванням зазначеного, LLaVa та Pixtral є важливими кандидатами для вирішення задачі "image-to-meteorology". Вони дозволяють використовувати готові потужні архітектури багатомодальних моделей, зводячи зусилля зі створення спеціалізованої системи до адаптації існуючих інструментів.

Подальше підвищення продуктивності, адаптивності і масштабованості пропонованого підходу забезпечується використанням мультиагентних систем (МАС) [21]. Як відомо, мультиагентний принцип дозволяє розподілити задачі між різними нейроагентами, кожен з яких спеціалізується на певному аспекті аналізу. Для вирішуваного завдання генерації метеоданих **мультиагентна система буде складатися з кількох взаємодіючих агентів, яких можна поділити на кілька категорій, наприклад:**

**агенти-збірники даних, що** отримують зображення хмар із різних джерел (супутників, БПЛА, літаків) та здійснюють їх попередню фільтрацію і обробку;

**агенти-аналітики, які** аналізують зображення та вилучають специфічні характеристики;

**агенти-прогнозисти** обчислюють метеодані на основі висновків агентів-аналітиків;  
**агенти-контекстуалізатори** уточнюють результати на основі поточної геолокації;  
**агенти-пояснювачі** формулюють результати у вигляді тексту або структурованого виходу.

Загальний принцип функціонування зазначеної МАС проілюстровано на рис. 4.

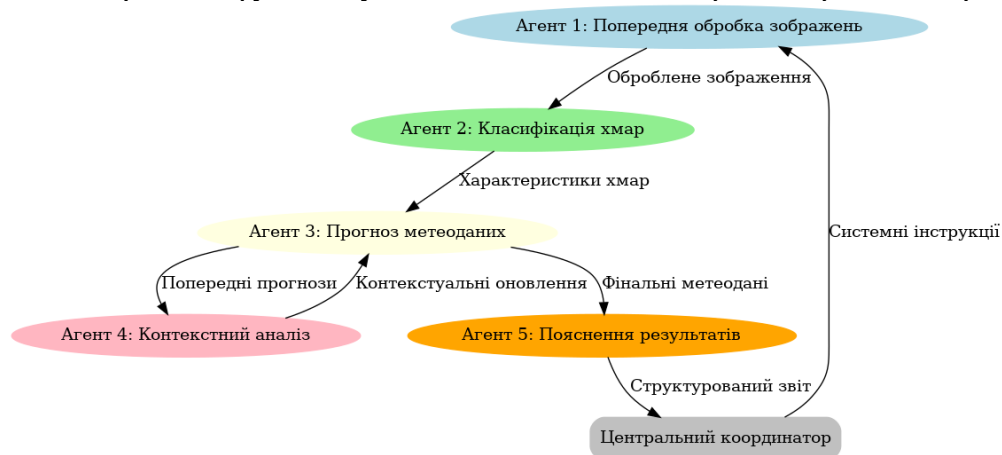


Рисунок 4. Принцип функціонування мультиагентної системи генерації метеоданих за зображеннями хмарних покрівів.

Розглянемо більш детально типові ролі агентів. Цілком очевидно, що агент з попередньої обробки зображень повинен займатися нормалізацією яскравості та контрасту зображення, виділенням хмарного середовища (сегментація з використанням, наприклад, U-Net), оцінкою ступеня суцільності хмар. Агент-аналітик з класифікації хмар далі використовує попередньо навчену нейромережну модель, щоб визначити тип хмар (перисті, шаруваті, купчасті тощо) і передає результати агенту-прогнозисту для подальших обчислень.

Агент-прогнозист розраховує параметри атмосферного середовища, такі як тиск, швидкість вітру і його напрямок, тип опадів. Для генерації текстових описів він може використовувати регресійні та трансформерні моделі або LLM, такі як LLaVa чи Pixtral. Агент з контекстного аналізу використовує геолокацію та історичні дані для уточнення прогнозу. Наприклад, якщо типові за структурою та кольором хмари в певному регіоні влітку зазвичай пов'язані з дощем, а не з градом, цей агент уточнить результати агента-прогнозиста. Для пояснення результатів прогнозу генеруються текстові або графічні коментарі у зручному для користувача форматі, наприклад: "Суцільні шаруваті хмари на висоті 2 км. Атмосферний тиск під хмарами - 1012 hPa. Опади: дощ."

Процес взаємодії агентів забезпечується шляхом їх комунікації один з одним. При цьому агенти обмінюються даними через шину повідомлень або загальний сервер, використовуючи формат даних типу JSON. Для координації функціонування може бути виділений централізований агент-координатор або розподілена система прийняття рішень, яка забезпечує узгодженість дій. Для кожного агента використовуються окремі спеціалізовані моделі (наприклад, MobileNet для класифікації, LLaVa для генерації текстів). Розгортання агентів може бути здійснено окремо у хмарному сервісі, на сервері або на борту БПЛА залежно від їх функцій.

Перевагами застосування мультиагентних систем є масштабованість та гнучкість, до них можна легко додавати інших агентів для нових задач (наприклад, прогнозування гроз чи снігопадів). За допомогою МАС завдання виконуються ефективніше, оскільки кожен агент відповідає лише за свою частину роботи й може використовувати різні моделі або алгоритми, такі як нейронні мережі, мовні моделі та методи машинного навчання для контекстного аналізу.

Описана концепція спостереження за хмарністю дозволить досліджувати глобальні кліматичні зміни, вивчати процеси утворення хмар, вплив вітрових течій на розподіл вологи

та динаміку опадів. У підсумку, зазначена ідея може стати не лише інструментом для прогнозування погоди, але й важливою частиною системи моніторингу кліматичних змін. Це відкриває широкий спектр можливостей як для практичного, так і для наукового застосування, особливо в умовах зростання уваги до кліматичних і метеорологічних питань у світі.

На завершення слід вказати на можливі напрями узагальнення пропонованого підходу прогнозування часових рядів через аналіз динаміки візуальних процесів, особливо для вирішення задач, пов'язаних із безпекою та управлінням ризиками. Сутність ідеї полягає у використанні візуального аналізу для послідовності зображень чи відеопотоку з метою моделювання часових рядів, що відображають розвиток процесів у часі. Наприклад, це може бути розповсюдження вогню навколо цистерни для прогнозування критичних подій, як-от момент вибуху. При такому сценарії на вхід системи надходить відеопотік або серія зображень, а вихідними даними є точка у часі до очікуваної критичної події, наприклад вибуху чи моменту початку евакуації персоналу, коли протидія пожежі вже втрачає ефективність.

Розробка відповідної системи штучного інтелекту включає кілька ключових етапів. Починається все зі збору й анотації даних, що охоплюють реальні записи пожеж або інших аварійних ситуацій. Для навчання можуть також використовуватися синтетичні дані, створені на основі фізичних моделей. Далі йде попередній аналіз відео, під час якого сегментація дозволяє виділяти критичні області. Це можна виконувати за допомогою моделей на основі U-Net, які ефективно розділяють зображення на необхідні зони. Наступний крок - аналіз динаміки, який передбачає оцінку характеристик, таких як площа поширення вогню, яскравість чи інтенсивність. Для цього можна використовувати нейромережні моделі ResNet, Vision Transformers та інші. Після вилучення ознак візуальної сцени вони подаються у модель часових рядів, яка прогнозує подальший розвиток події. Тут підходять такі моделі, як LSTM, GRU чи трансформери, які добре працюють із часовими залежностями. На основі навчання вказаних моделей можна побудувати систему, здатну прогнозувати час до настання критичної події. Описаний процес також включає регресійні розрахунки для точного визначення залишкового часу та класифікацію станів, наприклад, нормального, небезпечного чи критичного.

Для реалізації подібної системи в реальному часі необхідно мати обладнання, що здатне обробляти відеопотік без значних затримок. Це можна досягти за допомогою спеціалізованих обчислювальних платформ, таких як NVIDIA TensorRT або OpenVINO. Постійний моніторинг дозволяє перераховувати залишковий час до критичної події і передавати попередження персоналу, що значно підвищує безпеку. Можливості застосування даного підходу надзвичайно широкі. У промисловості система може бути використана для моніторингу пожеж на об'єктах із небезпечними речовинами. Вона також може прогнозувати руйнування інфраструктурних об'єктів, таких як мости чи дамби, шляхом аналізу динаміки тріщин чи вібрацій, а у рятувальних службах – спрогнозувати час для евакуації під час природних катастроф. У медичній сфері подібний аналіз може застосовуватися для прогнозування кризових станів пацієнтів за динамікою зміни на моніторах фізіологічних параметрів, таких як серцебиття чи дихання.

Звісно, розробка подібних систем стикається з викликами, такими як дефіцит якісних даних, необхідність високої точності прогнозів та забезпечення роботи в реальному часі. Разом з тим, вирішення цих проблем відкриє нові перспективи для впровадження штучного інтелекту у прогнозування кризових ситуацій, що базуються на динаміці візуальних даних.

#### Список посилань.

1. *Гетьман, Ю. А. Солод, М. А. Держевецька. Дослідження застосування штучних нейронних мереж для підвищення точності прогнозування температури навколишнього середовища. Вісник Херсонського національного технічного університету, № 2 (89), 2024. [Онлайн]. Доступно: [https://journals.kntu.kherson.ua/index.php/visnyk\\_kntu/article/view/638](https://journals.kntu.kherson.ua/index.php/visnyk_kntu/article/view/638).*
2. *Малярчук, Р. В. Класифікація хмар на зображеннях за допомогою глибоких нейронних мереж : дипломна робота ... бакалавра : 124 Системний аналіз / Малярчук Роман Васильович. - Київ, Національний*



технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", 2020. 106 с.

3. Ye, L.; Cao, Z.; Xiao, Y.; Li, W. Ground-based cloud image categorization using deep convolutional visual features. In *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015*; pp. 4808–4812.
4. Shi, C.; Wang, C.; Wang, Y.; Xiao, B. Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geosci. Remote Sens. Lett.* 2017, 14, 816–820.
5. Zhang, J.; Liu, P.; Zhang, F.; Song, Q. CloudNet: Ground-based cloud classification with deep convolutional neural network. *Geophys. Res. Lett.* 2018, 45, 8665–8672.
6. Wang, M.; Zhou, S.; Yang, Z.; Liu, Z. CloudA: A Ground-Based Cloud Classification Method with a Convolutional Neural Network. *J. Atmos. Ocean. Technol.* 2020, 37, pp. 1661–1668.
7. Liu, S.; Duan, L.; Zhang, Z.; Cao, X.; Durrani, T.S. Multimodal ground-based remote sensing cloud classification via learning heterogeneous deep features. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, pp. 7790–7800.
8. Liu, S.; Li, M.; Zhang, Z.; Xiao, B.; Cao, X. Multimodal ground-based cloud classification using joint fusion convolutional neural network. *Remote Sens.* 2018, 10, 822.
9. Liu, S.; Li, M.; Zhang, Z.; Xiao, B.; Durrani, T.S. Multi-evidence and multi-modal fusion network for ground-based cloud recognition. *Remote Sens.* 2020, 12, 464.
10. Lv, Q.; Li, Q.; Chen, K.; Lu, Y.; Wang, L. Classification of Ground-Based Cloud Images by Contrastive Self-Supervised Learning. *Remote Sens.* 2022, 14, 5821. DOI: 10.3390/rs14225821.
11. Yousaf, R.; Rehman, H.Z.U.; Khan, K.; Khan, Z.H.; Fazil, A.; Mahmood, Z.; Qaisar, S.M.; Siddiqui, A.J. Satellite Imagery-Based Cloud Classification Using Deep Learning. *Remote Sens.* 2023, 15, 5597. DOI: 10.3390/rs15235597.
12. Nilay S. Kapadia, Urmil Parikh, Dipti P. Rana. Weather Forecasting using Satellite Image Processing and Artificial Neural Networks. *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 11, November 2016, pp. 1069 – 1074.
13. Ігор Кібальчич. Штучний інтелект та нейронні мережі у метеорологічній практиці. 2023. [Онлайн]. Доступно: <https://www.meteoprog.com/ua/news/623350-stucnii-intelekt-ta-neironni-merezi-u-meteorologichniy-praktici.html>.
14. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Pavlo Kovalov, Pavlo Borodych, Serhii Shevchenko, Oleksandr Chernikov, Serhii Vazhynskyi, Oleg Bogatov, Kirill Khrustalev. Improvement of the object recognition model on aerophotos using deep conventional neural network. // *Eastern-European Journal of Enterprise Technologies.* - 2021, Vol. 5, No. 2 (113). Pp. 6 – 21. DOI: 10.15587/1729-4061.2021.243094.
15. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Vasyl Melkin, Oleh Biloborodov, Mykola Samoilenko, Olena Kravchenko, Galina Kalinichenko, Anton Rohovyi, Mykhaylo Soloshchuk. Improvement of the model for detecting objects on aerial photos and video in unmanned aerial systems. // *Eastern-European Journal of Enterprise Technologies.* - Vol. 1, No. 9(115). – 2022.- Pp.24 - 34. DOI: 10.15587/1729-4061.2022.252876.
16. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Stella Gornostal, Sergey Rudakov, Serhii Shevchenko, Oleksandr Chernikov, Nadiia Kolpachenko, Volodymyr Timofeyev, Roman Artiukh. Construction of an advanced method for recognizing monitored objects by a convolutional neural network using a discrete wavelet transform. // *Eastern-European Journal of Enterprise Technologies*, Vol. 4 No. 9(112) (2021): Information and controlling system, Pp. 65 - 77. DOI: 10.15587/1729-4061.2021.238601.
17. Andrii Zinchenko, Vadym Slyusar, Natalia Korolyuk, Elena Korshets. The Method of Open Space Selection of Signals for Radcom Systems. // *3rd International Conference on Advanced Information and Communications Technologies (AICT'2019)*. – 2 - 6 July 2019. - Lviv Polytechnic National University, Lviv, Ukraine. - Pp. 404-408. - DOI: 10.1109/AIACT.2019.8847826.
18. Слюсар В.І., Слюсарь І.І. Класифікація зимового та літнього сезонів з борта БПЛА. // *Нейромережні технології та їх застосування НМТіЗ-2023: збірник наукових праць XXII-ї Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2023», 13-14 грудня 2023 року, Краматорськ-Тернопіль, Україна.* - С. 134 -142.
19. Tsang S. Review: MobileNetV2 – Light Weight Model (Image Classification). URL: <https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c>. (дата звернення: 27.10.2023).
20. ImageNet. URL: <https://www.image-net.org>. (дата звернення: 27.10.2023).
21. Слюсар В.І. Локальні великі мовні моделі для обробки конфіденційної інформації. // *Озброєння та військова техніка*, № 4(44), 2024. – С. 79 – 91. DOI: 1034169/2414-0651.2024.4(44).79-91.

## ВІЗУАЛЬНА ІНТЕРПРЕТАЦІЯ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ КІБЕРБУЛІНГУ У ЦИФРОВИХ ТЕКСТАХ

**Анотація:** Запропоновано метод візуальної інтерпретації нейромережевого виявлення кібербулінгу в цифрових текстах, що дозволяє інтерпретувати рішення моделі щодо типів кібербулінгу. Метод базується на використанні моделі BERT для мультилейблової класифікації та інтерпретаційної моделі LIME, яка візуалізує вплив слів на рішення моделі. Метод забезпечує три формати інтерпретації: кольорову палітру, діаграми локальної та загальної важливості слів. Експерименти підтвердили, що розроблений підхід забезпечує зрозуміле пояснення рішень штучного інтелекту щодо виявлених типів кібербулінгу..

**Ключові слова:** кібербулінг, інтерпретація результатів, нейронні мережі, BERT, LIME.

**Abstract:** A method for explaining the results of neural network detection of cyberbullying in digital texts is proposed, which allows interpreting the model's decisions regarding the types of cyberbullying. The method is based on the use of the BERT model for multi-label classification and the LIME interpretation model, which visualizes the influence of words on the model's decisions. The method provides three interpretation formats: a color palette, diagrams of local and global word importance. Experiments have confirmed that the developed approach provides a clear explanation of artificial intelligence decisions regarding the detected types of cyberbullying..

**Keywords:** propaganda objects, propaganda techniques, propaganda detection, natural language processing

### Постановка проблеми

Проблема кібербулінгу стає дедалі актуальнішою через зростання кількості користувачів соціальних мереж, особливо серед молоді, що збільшує попит на системи нейромережевого виявлення кібербулінгу в цифрових текстах [1,2]. Завдяки прогресу у використанні моделей трансформерів, зокрема BERT, стало можливим ефективно виявляти та класифікувати типи кібербулінгу [3]. Однак складність інтерпретації таких моделей викликає сумніви щодо їх використання у чутливих контекстах. Тому інтерпретація рішень є ключовою для забезпечення довіри та прозорості. У роботі запропоновано метод пояснення рішень моделі щодо виявлених типів кібербулінгу, таких як дискримінація за віком, етнічністю чи гендером.

### Аналіз останніх публікацій

Проблема нейромережевого виявлення кібербулінгу є надзвичайно актуальною через його руйнівний вплив на психічне здоров'я, особливо підлітків та молоді. Сучасні методи ґрунтуються на технологіях обробки природної мови, що дозволяють аналізувати цифрові тексти для виявлення та класифікації різних форм кібербулінгу [4].

У дослідженні [5] розглядається задача нейромережевого виявлення кібербулінгу. Серед протестованих моделей, таких як Random Forest, XgBoost, Naive Bayes, SVM, CNN, RNN та BERT, остання продемонструвала найвищу ефективність, досягнувши 88,8% точності у бінарній класифікації та 86,6% у мультилейбловій.

Автори роботи [6] розробили новий підхід до виявлення кібербулінгу, протестувавши SVM, Naive Bayes і Logistic Regression у поєднанні з різними методами обробки тексту. Було доведено, що аналіз настроїв, N-грам, TF-IDF та визначення ненормативної лексики суттєво покращують точність, дозволяючи досягти 75,17% у задачі класифікації.

Інші автори зосередили увагу на інтерпретації результатів. Наприклад, у [7] представлено модель BiLSTM-LIME для багатокласової класифікації кібербулінгу в цифрових текстах Twitter. Використання LIME забезпечило високу якість пояснень, акцентуючи увагу на токенах, які вплинули на рішення.

Дослідження [8] запропонувало ансамбль BERT та SVM з налаштуванням параметрів для багатокласової класифікації кібербулінгу у соціальних медіа. Модель показала точність 90% на тестових даних, перевершивши альтернативні підходи. Для пояснення прогнозів було використано техніку SHAP, яка надала детальний аналіз значущості ознак.

Аналіз публікацій свідчить, що виявлення кібербулінгу в цифрових текстах є важливим і багатограним завданням, яке активно досліджується завдяки значному впливу цієї проблеми на суспільство. Використання сучасних моделей обробки природної мови, особливо архітектур трансформерів, таких як BERT, демонструє високу ефективність у задачах як бінарної, так і мультилейблової класифікації кібербулінгу. Зокрема, BERT стабільно перевершує інші підходи, досягаючи точності понад 85% у більшості експериментів. Що стосується інтерпретації, роботи підкреслюють важливість забезпечення прозорості моделей. Методи, такі як LIME та SHAP, дозволяють не лише пояснити рішення моделі, але й зробити її застосування більш зрозумілим для користувачів. Це особливо важливо у соціально значущих контекстах, де прозорість рішень безпосередньо впливає на довіру до систем штучного інтелекту.

### Мета роботи та постановка завдань

Мета роботи полягає в розробці методу для візуальної інтерпретації результатів нейромережевого виявлення кібербулінгу в цифрових текстах, спрямованого на пояснення рішень моделі штучного інтелекту стосовно визначених типів кібербулінгу. Запропонований метод повинен забезпечувати зрозумілу інтерпретацію, яка дозволяє людині аналізувати текстові ознаки, що вплинули на рішення нейромережевої моделі щодо ідентифікації типів кібербулінгу.

### Виклад основного матеріалу

Метод інтерпретації результатів нейромережевого виявлення кібербулінгу в цифрових текстах передбачає створення візуального пояснення рішень моделі штучного інтелекту щодо визначених типів кібербулінгу [9]. Схематичне представлення цього методу наведено на рисунку 1.

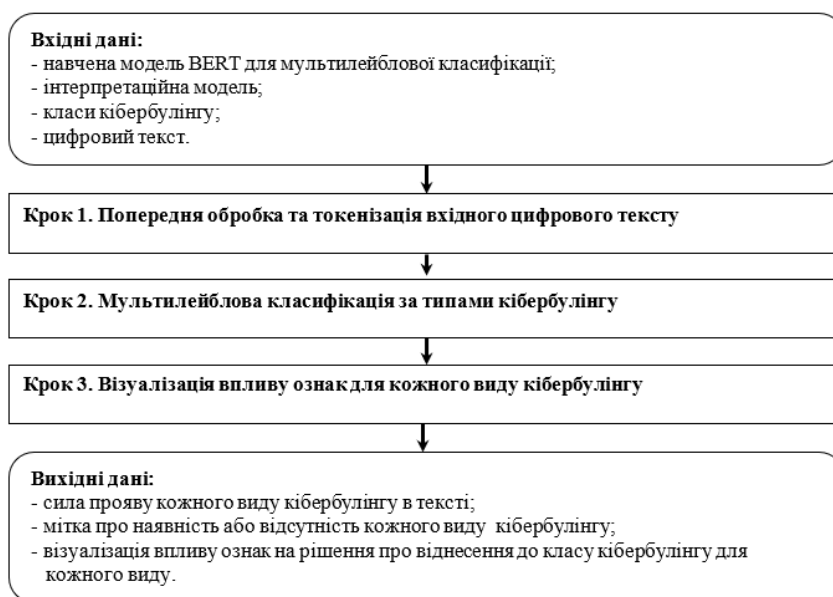


Рисунок 1. Схема методу візуальної інтерпретації нейромережевого виявлення кібербулінгу в цифрових текстах

Вхідними даними цього етапу є навчена модель трансформерної архітектури для мультилейблової класифікації, здатна визначати різні типи кібербулінгу, такі як віковий, етнічний, гендерний, релігійний та узагальнений тип, що охоплює інші види кібербулінгу. Також використовуються інтерпретаційні моделі, які пояснюють вплив окремих слів чи фраз

на результати класифікації. Вхідний текст аналізується на наявність ознак кібербулінгу, після чого результати піддаються інтерпретації.

Першим кроком є токенизація тексту, після чого текстові елементи перетворюються в числові послідовності для подальшої обробки нейромережевою моделлю.

Другий крок полягає у прогнозуванні ймовірностей належності тексту до кожного з типів кібербулінгу, оцінюючи наявність ознак, таких як вікові, етнічні чи гендерні характеристики.

Третім етапом є пояснення та візуалізація результатів класифікації за допомогою інтерпретаційної моделі, яка виявляє вплив окремих слів або фраз на ідентифікацію ознак кібербулінгу. Для мультілейбрової класифікації часто застосовуються такі інтерпретаційні методи, як [10]: Local Interpretable Model-agnostic Explanations (LIME), який генерує локальні пояснення для кожного передбачення, демонструючи, які слова найбільше вплинули на результат; SHapley Additive exPlanations (SHAP), що базується на теорії ігор і обчислює внесок кожного слова у передбачення, враховуючи взаємодію між ознаками; Transformers Interpret, інтерпретаційна бібліотека, розроблена спеціально для моделей на основі трансформерів, таких як BERT, GPT, RoBERTa та інші моделі з бібліотеки Hugging Face; методи, що використовують Attention, які аналізують ваги уваги трансформерів (наприклад, у моделі BERT) для розуміння важливості окремих слів чи фраз у процесі прийняття рішень моделлю.

Вихідними даними є інтенсивність прояву кожного типу кібербулінгу в тексті, виражена через ймовірності, які демонструють ступінь наявності ознак для кожного класу кібербулінгу. Для кожного класу визначається мітка, що вказує на наявність або відсутність ознак, представлених числовими значеннями, які відображають ймовірність прояву кібербулінгу за кожним типом. Крім того, метод забезпечує візуалізацію впливу конкретних ознак на прийняте рішення про належність тексту до певного класу кібербулінгу, де важливі слова підсвічуються відповідно до їх значущості для кожного з класів.

Таким чином наведений метод візуальної інтерпретації результатів нейромережевого виявлення кібербулінгу сприятиме кращому розумінню та поясненню рішень, ухвалених моделлю щодо мультілейбрової класифікації цифрових текстів та визначених типів кібербулінгу.

Для навчання моделі BERT [11], яка застосовується на кроці 2 методу візуальної інтерпретації нейромережевого виявлення кібербулінгу (рисунок 1), використовувався датасет «Cyberbullying Classification» [12]. Цей датасет містить текстові повідомлення з мітками, що визначають належність кожного повідомлення до одного з класів: Age, Ethnicity, Gender, Religion, Other type of cyberbullying, Not cyberbullying.

Для навчання моделі BERT мультілейбровій класифікації був видалений клас «Not cyberbullying» з датасету «Cyberbullying Classification», оскільки він не використовувався в навчанні. Крім того, клас «Other type of cyberbullying» був збільшений за допомогою методики SMOTE-балансування, що дозволило створити синтетичні зразки. Завдяки цьому попередньому етапу обробки даних був отриманий збалансований набір для навчання моделі BERT для завдання мультілейбрової класифікації типів кібербулінгу в текстовому контенті.

Для оцінки ефективності методу візуальної інтерпретації нейромережевого виявлення кібербулінгу в цифрових текстах використовувалося середовище Google Colab. Модель BERT була навчена для класифікації таких типів кібербулінгу, як віковий, гендерний, релігійний, етнічний, а також окремо для типу «інші кібербулінги».

Показники макрометрик навченої моделі BERT для мультілейбрової класифікації типів кібербулінгу становлять: Accuracy 0.956478, Precision 0.963677, Recall 0.956478 та F1 Score 0.960019. Ці значення свідчать про високу ефективність моделі у виявленні різних видів кібербулінгу в текстовому контенті.

Для дослідження був використаний англomовний цифровий текст, який було проаналізовано для виявлення різних типів кібербулінгу за допомогою навченої моделі BERT. Модель BERT виявила ймовірності наявності різних видів кібербулінгу в цифровому тексті,

зокрема віковий кібербулінг – 0.06%, етнічний – 0.08%, гендерний – 0.10%, інший тип – 0.09%, та релігійний кібербулінг – 99.86%

Застосування моделі LIME для візуальної інтерпретації нейромережевого виявлення кібербулінгу за допомогою моделі BERT для мультисловової класифікації типів кібербулінгу в цифровому тексті дозволило отримати візуальні результати інтерпретації виявлених типів кібербулінгу, використовуючи абсолютні значення ваг, що зображені на рисунку 2. Для пояснення прийнятих рішень моделлю BERT слова в цифровому тексті виділяються різними кольорами: найбільш яскравий колір вказує на найбільшу вагу слова, що означає його найбільший вплив на результат, а найсвітліший – на найменший.

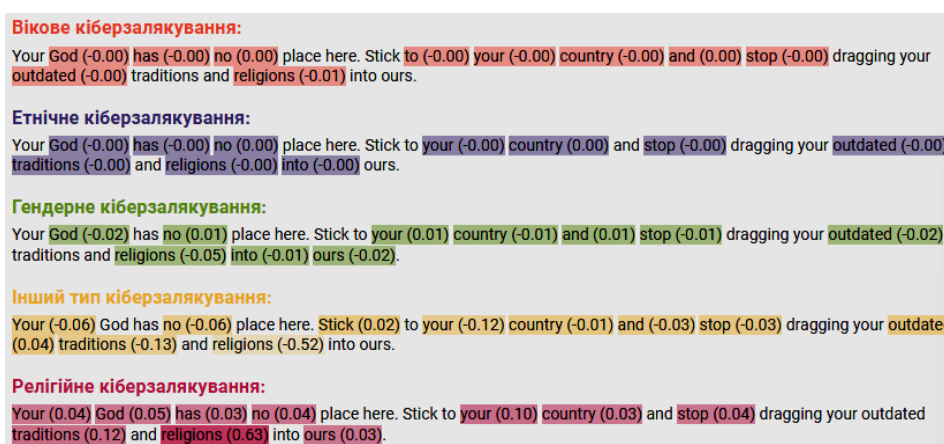


Рисунок 2. Абсолютне значення ваги для визначення яскравості кольору з метою інтерпретації результатів виявлення різних типів кібербулінгу в цифровому тексті

Як видно з рисунку 2, слова з додатними та від'ємними значеннями виділяються однаковою яскравістю. В цьому випадку для визначення яскравості використовується абсолютне значення ваги, що призводить до однакової яскравості для від'ємних та додатних значень. Від'ємні значення ваги зменшують ймовірність певного класу, тоді як додатні значення збільшують її, але обидва типи мають однаковий вплив на прийняте моделлю рішення. Для LIME важливо не лише показати силу впливу слова, а й його напрямок (позитивний або негативний). Тому реалізовано підхід, де від'ємні значення мають менш яскравий колір і окремий відтінок для додатних та від'ємних значень. Результати такої візуалізації подано на рисунку 3. Використання різних кольорів для додатних і від'ємних значень є важливим, оскільки від'ємні ваги зменшують ймовірність певного класу, а додатні – збільшують. Без цієї відмінності, однакові інтенсивності різних знаків можуть бути сприйняті як рівнозначні, що може призвести до неправильного розуміння результатів.



Рисунок 3. Підхід для інтерпретації результатів виявлення типів кібербулінгу з урахуванням негативного чи позитивного типу впливу кінцевий результат

Додатково були створені діаграми для графічної інтерпретації впливу окремих слів цифрового тексту на ймовірність віднесення цього тексту до конкретного типу кібербулінгу (рисунок 4).

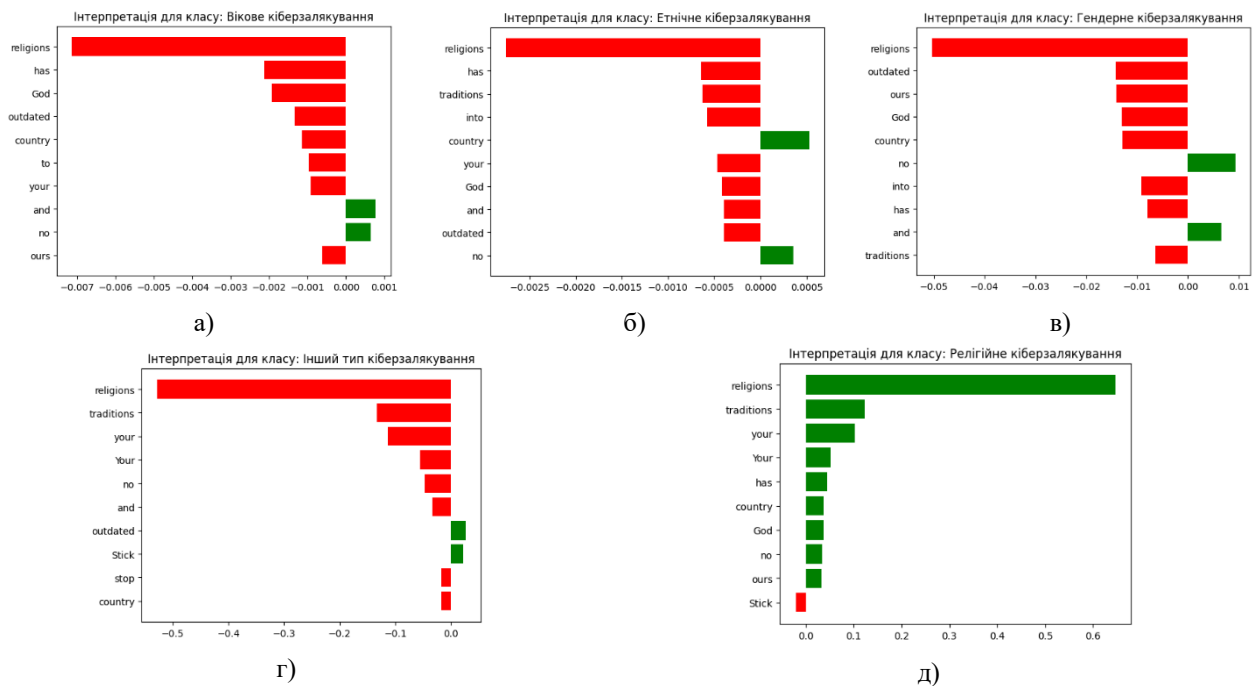


Рисунок 4. Графічна інтерпретації впливу окремих слів цифрового тексту на результат

Діаграми показують, як модель оцінює вагу кожного слова в цифровому тексті, залежно від його внеску в прийняте рішення. Вплив слів відображено у вигляді горизонтальних стовпців, довжина яких показує величину впливу (ваги), а колір – напрямок цього впливу. Червоні стовпці вказують на негативний вплив слів, що зменшують ймовірність віднесення тексту до певного класу, тоді як зелені стовпці означають позитивний вплив, який збільшує ймовірність цього віднесення. Величина впливу вимірюється числовим значенням, що відображається на горизонтальній осі графіка.

Також було обчислено середнє значення важливості кожного слова для всіх класів, що дозволяє оцінити загальний вплив кожного слова без прив'язки до конкретного типу кібербулінгу. Результати обчислень представлені у вигляді діаграми (рисунок 5).

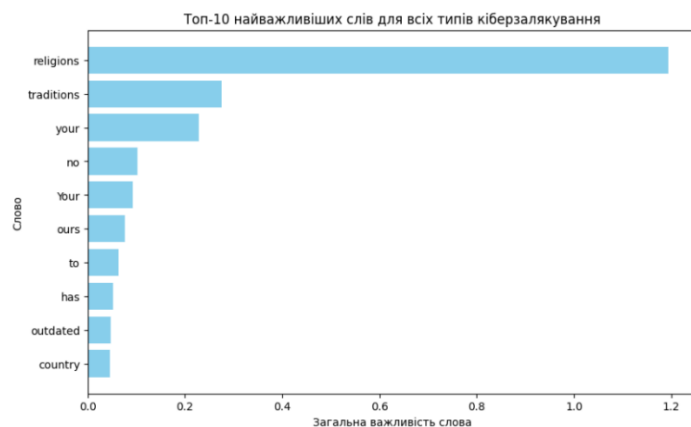


Рисунок 5. Середні значення важливості топ-10 слів для всіх класів

Обчислення загального впливу слів на результати моделі для всіх типів кібербулінгу є важливим для розуміння роботи моделі та її рішень. Аналіз здійснюється через агрегацію ваг слів, які модель оцінює для кожного класу. Використовується модуль ваги, що означає абсолютну величину впливу слова без урахування його позитивного чи негативного значення. Цей підхід дає можливість виявити слова, які модель вважає важливими незалежно від конкретного типу кібербулінгу. Наприклад, слова, що стосуються різних типів кібербулінгу, можуть мати високі ваги для кількох класів. Якщо слово має високий загальний вплив, це може свідчити про його універсальну роль у контексті кібербулінгу. Наприклад, слова, що вказують на етнічну приналежність або релігію, можуть мати великий вплив на кілька класів, таких як «етнічний кібербулінг» і «релігійний кібербулінг», що вказує на потенційну крос-модальність ознак, які модель використовує для прийняття рішень. Якщо ж слово має вплив лише на один клас, це підкреслює його специфічність і може вказувати на унікальні мовні патерни для цього виду кібербулінгу.

Отже, візуальні інтерпретації результатів нейромережевого виявлення кібербулінгу в цифрових текстах дозволяють оцінити, чи модель використовує релевантні ознаки для ухвалення рішень, чи її поведінка зумовлена випадковими чи нерелевантними факторами. Наприклад, якщо в тексті з'являються слова, що не мають змістового зв'язку з віковим кібербулінгом, але мають значний вплив, це може свідчити про наявність помилки або упередження в моделі.

## ВИСНОВКИ

У роботі запропоновано метод візуальної інтерпретації нейромережевого виявлення кібербулінгу в цифрових текстах, призначений для пояснення рішень нейромережевої моделі щодо типів кібербулінгу, виявлених у текстах. Метод є оригінальним, оскільки здійснює інтерпретацію результатів для кожного типу кібербулінгу окремо, що досягається використанням мультитейблових класифікаторів нейромережевої архітектури трансформерів і інтерпретаційної моделі машинного навчання.

Завдяки використанню навченої нейромережевої моделі BERT для мультитейблової класифікації типів кібербулінгу в цифровому тексті, модель виявляє різні типи кібербулінгу з указанням відсотка наявності кожного з них. Згідно з розробленим методом, для візуальної інтерпретації результатів виявлення кібербулінгу використано підхід, що базується на моделі машинного навчання LIME для локальної інтерпретованості, що дозволяє візуалізувати вплив кожного окремого слова на рішення моделі щодо належності тексту до певних типів кібербулінгу.

Метод забезпечує три способи візуальної інтерпретації нейромережевого виявлення кібербулінгу в цифрових текстах: за кольоровою палітрою, за діаграмами локальної важливості слів і за діаграмами загальної важливості слів. Інтерпретація результатів за кольоровою палітрою ґрунтується на використанні абсолютного значення ваги для визначення яскравості кольору, де найбільш яскравий колір вказує на найбільший вплив слова на прийняте рішення моделі, а найменш яскравий – на найменший вплив, незалежно від того, чи був він позитивним чи негативним. Проте, для повної інтерпретації необхідно також розуміти напрямок впливу, оскільки від'ємні ваги зменшують ймовірність певного класу, а додатні – збільшують її. Тому реалізовано інтерпретацію рішень моделі BERT з урахуванням напрямку впливу.

Візуальна інтерпретація результатів за діаграмами локальної важливості слів демонструє, як кожне слово впливає на ймовірність віднесення тексту до конкретного типу кібербулінгу, дозволяючи побачити, як модель оцінює вагу кожного слова, залежно від його впливу на прийняте рішення. Інтерпретація результатів за діаграмами загальної важливості слів показує 10 слів, які модель вважає важливими для визначення типу кібербулінгу, незалежно від конкретного класу.

Результати експериментів свідчать, що запропонований метод забезпечує візуальну інтерпретацію рішень щодо нейромережевого виявлення кібербулінгу на рівні, який дозволяє людині зрозуміти, які ознаки тексту вплинули на прийняття рішень штучним інтелектом. Розроблений метод інтерпретації виявлення кібербулінгу у цифрових текстах належить до категорії засобів візуальної аналітики рішень штучного інтелекту, що є необхідним для забезпечення етичності, прозорості та довіри до таких систем штучного інтелекту в

суспільстві, особливо коли йдеться про чутливі питання, як кібербулінг. Дослідження підкреслює важливість не тільки точності моделей, але й їхньої пояснюваності, що є ключовим для побудови довіри до систем штучного інтелекту.

#### Список посилань.

1. Собко О.В. Виявлення та класифікація кіберзалякувань у цифрових текстах засобами штучного інтелекту / О.В. Собко // Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах». – 2024. – № 4. – С. 143–152.
2. Krak I. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network / I. Krak, O. Zalutska, M. Molchanova, O. Mazurets, R. Bahrii, O. Sobko, O. Barmak // CEUR Workshop Proceedings. – 2024. – Vol. 3688. – С. 16–28.
3. Собко О.В. Метод інтелектуального виявлення та класифікації кіберзалякувань у текстовому контенті / О.В. Собко // Інформаційні управляючі системи та технології ІУСТ-ОДЕСА-2024: матеріали XII Міжнар. наук.-практ. конф.– Одеса, 2024. – С. 262–265.
4. Молчанова М.О. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі / М.О. Молчанова, О.В. Мазурець, О.В. Собко, В.І. Кліменко, В.І. Андрощук // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – № 2 (333). – С. 200–206.
5. Sen M. From Tweets to Insights: BERT-Enhanced Models for Cyberbullying Detection / M. Sen, J. Masih, R. Rajasekaran // 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS): Proc. – 2024. – С. 1289–1293.
6. Abood M.M. Explainable Multimodal Deep Learning Model for Cyberbullying Detection (EMDL-CBD) / M.M. Abood, M.A. Al-Bayati // Journal Port Science Research. – 2024. – Vol. 7, № 3.
7. Nuthalapati P. Cyberbullying Detection: A Comparative Study of Classification Algorithms [Електронний ресурс] – Режим доступу: <https://www.authorea.com/doi/full/10.22541/au.170664263.38254624> (дата звернення: 17.17.2024).
8. Perera A. Cyberbullying Detection System on Social Media Using Supervised Machine Learning / A. Perera, P. Fernando // Procedia Computer Science. – 2024. – Vol. 239. – С. 506–516.
9. Molchanova M. Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP / M. Molchanova, O. Mazurets, O. Sobko, I. Boiarchuk // Scientific Achievements and Innovations as a Way to Success: Proc. XXI Int. Scientific and Practical Conf., May 1–3, 2024, Vilnius, Lithuania. – Vilnius, 2024. – С. 73–77.
10. Kiefer S. CaSE: Explaining Text Classifications by Fusion of Local Surrogate Explanation Models with Contextual and Semantic Knowledge / S. Kiefer // Information Fusion. – 2022. – Vol. 77. – С. 184–195.
11. Alissa S. Text Simplification Using Transformer and BERT / S. Alissa, M. Wald // Computers, Materials & Continua. – 2023. – Vol. 75, № 2. – С. 3479–3495.
12. Cyberbullying Classification Dataset [Електронний ресурс]. – Kaggle. – Режим доступу: <https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification> (дата звернення: 17.17.2024).



Чабан О.Р., Манзюк Е.А. (Хмельницький національний університет, м. Хмельницький, Україна)

## ПІДХІД ДО ІНТЕГРУВАННЯ ЕКСПЕРТНИХ ЗНАТЬ В МОДЕЛЬ U-NET ДЛЯ СЕГМЕНТУВАННЯ ЗОБРАЖЕНЬ МРТ СЕРЦЯ

**Анотація:** У статті запропоновано новий підхід до інтегрування експертних знань медичних фахівців у нейронну мережу архітектури U-Net для сегментування зображень МРТ серця. Запропонований підхід передбачає використання механізмів уваги, керованих експертними анотаціями, та впровадження анатомічних обмежень через спеціалізовані функції втрат. Це дає змогу покращити здатність мережі створювати анатомічно точні та клінічно значущі сегментовані маски на зображенні МРТ серця. Підхід включає декілька послідовних етапів, зокрема, аугментацію даних з експертними анотаціями, проєктування механізму уваги та додання функцій втрат із врахуванням анатомічних обмежень. Експериментальні результати за набором даних Automated Cardiac Diagnosis Challenge у задачі сегментування продемонстрували покращення проти базової моделі U-Net за низкою метрик сегментації.

**Ключові слова:** інтеграція експертних знань, нейронна мережа, U-Net, сегментація, МРТ серця, механізм уваги, анатомічні обмеження, функції втрат, аугментація даних, Automated Cardiac Diagnosis Challenge.

**Abstract:** This study presents a novel method for integrating expert domain knowledge from healthcare professionals into a U-Net-type neural network for cardiac MRI segmentation. By incorporating attention mechanisms guided by expert annotations and enforcing anatomical constraints through specialized loss functions, our approach enhances the network's ability to produce anatomically accurate and clinically meaningful segmentations. The proposed method is detailed in several sequential steps, including data augmentation with expert annotations, the design of an attention mechanism, and the incorporation of anatomical constraint loss functions. Experimental results on the Automated Cardiac Diagnosis Challenge dataset demonstrate significant improvements over the baseline U-Net model across several segmentation metrics.

**Keywords:** integration of expert knowledge, neural network, U-Net, segmentation, cardiac MRI, attention mechanism, anatomical constraints, loss functions, data augmentation, Automated Cardiac Diagnosis Challenge.

### Вступ.

Магнітно-резонансна томографія (МРТ) серця є основним інструментом у процесі діагностування та лікування серцево-судинних захворювань. Вона забезпечує деталізоване зображення структур серця без іонізуючого випромінювання. Точне сегментування зображень МРТ серця є важливою для кількісного оцінювання функції серця, аналізу життєздатності міокарда та планування втручань [1]. Згорткові нейронні мережі (CNN), зокрема архітектура U-Net [2], показали високі результати в задачах сегментування медичних зображень [3]. Однак ці моделі часто працюють без явного інтегрування експертних знань медичних фахівців [4, 5], що може обмежувати їхню здатність створювати анатомічно достовірні та клінічно правдиві сегментовані маски.

Медичні фахівці володіють експертним глибоким знанням щодо анатомії серця, патологій та клінічного значення певних ознак на зображеннях. Інтегрування цих знань у моделі глибокого навчання може значно підвищити їхню продуктивність, особливо у випадках зі складною анатомією чи артефактами [6, 7]. Сучасні підходи, як от механізми уваги [8] та функції втрат з анатомічними обмеженнями [9], дають перспективи для впровадження експертного досвіду в нейронні мережі [10].

У цій роботі запропоновано новий підхід до інтегрування експертних знань у нейронну мережу типу U-Net для сегментування зображень МРТ серця. Наш підхід використовує механізми уваги, керовані експертними анотаціями, для фокусування мережі на клінічно важливих областях, а також включає анатомічні обмеження у функції втрат для забезпечення реалістичних сегментацій.

## Запропонований підхід.

Запропонований підхід до інтегрування експертних знань у нейронну мережу включає кілька послідовних етапів, що покликані забезпечити результативну та успішну інтеграцію експертних знань.

**Блок 1.** Аугментація даних з експертними анотаціями

*Крок 1.1.* Збір експертних анотацій

Медичні фахівці надають анотації, що виділяють області інтересу (ROI) на зображеннях МРТ серця, як от лівий шлуночок (LV), правий шлуночок (RV) і міокард, включно з областями інтересу (ROI) МРТ серця, що схильні до помилок сегментування через патології або артефакти.

*Крок 1.2.* Аугментація навчальних даних

Експертні анотації використовуються для створення додаткових каналів у вхідних даних, що дає змогу мережі навчатися як за вхідними зображеннями, так і за областями МРТ, що виділені експертами.

**Блок 2.** Механізм уваги за експертним знанням

*Крок 2.1.* Інтеграція механізмів уваги

Архітектура U-Net модифікована через додавання порогів уваги (attention gates) у з'єднаннях із пропуском (skip connections). Ці пороги отримують мапи ознак з енкодера та ROI-дані від експертів, формуючи коефіцієнти уваги, що виділяють важливі області.

Нехай  $X_i$  – вхідна мапа ознак на 1-ому шарі, а  $G$  – сигнал-напряв від експертних анотацій. Коефіцієнт уваги  $\alpha$  обчислюється за формулою:

$$\alpha_i = \sigma(W^T [X_i, G] + b),$$

де  $W$  і  $b$  – ваги, значення яких визначаються під час навчання нейронної мережі,  $[\cdot, \cdot]$  – конкатенація, а  $\sigma$  – сигмоїдна активаційна функція.

*Крок 2.2.* Модуляція мап ознак

Коефіцієнти уваги модулюють мапи ознак:

$$\tilde{X}_i = \alpha_i \square X_i,$$

де  $\square$  – покомпонентне множення.

Наведена вище модуляція мап ознак має на меті фокусувати увагу мережі на регіонах, які експерти вважають важливими.

**Блок 3.** Впровадження функцій втрат з анатомічними обмеженнями

*Крок 3.1.* Визначення анатомічних співвідношень

На основі експертних знань подамо такі анатомічні обмеження:

1. Міокард мусить повністю охоплювати порожнину LV.
2. Порожнина RV має бути прилеглою до міокарда.
3. Неприпустиме перекриття структур, які не є сусідніми.

*Крок 3.2.* Функції втрат для анатомічних обмежень

Нижче формалізуємо функції втрат, що спроектовані виключно для усунення поданих вище обмежень:

– функція втрат охоплення,  $L_{\text{enc}}$  :

$$L_{\text{enc}} = \max(0, \text{Area}(LV \setminus \text{Myocardium}));$$

– функція втрат суміжності,  $L_{\text{adj}}$  :

$$L_{\text{adj}} = \text{Distance}(RV, \text{Myocardium});$$

– функція втрат перекриття,  $L_{\text{overlap}}$  :

$$L_{\text{overlap}} = \text{Area}(\text{Overlap of non-adjacent structures}).$$

*Крок 3.3.* Загальна функція втрат для анатомічних обмежень

Загальна функція втрат визначено в такий спосіб:

$$L_{\text{anatomy}} = \lambda_{\text{enc}} L_{\text{enc}} + \lambda_{\text{adj}} L_{\text{adj}} + \lambda_{\text{overlap}} L_{\text{overlap}},$$

де  $\lambda_{enc}$ ,  $\lambda_{adj}$ ,  $\lambda_{overlap}$  – вагові коефіцієнти.

У роботі було впроваджено додаткові задачі, що враховують експертні знання.

а) Визначення анатомічних орієнтирів: Прогнозування ключових точок анатомії.

б) Класифікація фаз серця: Визначення фаз серцевого циклу (систола, діастола).

### Опис даних.

Для оцінювання запропонованого підходу було використано набір даних Automated Cardiac Diagnosis Challenge (ACDC) [1]. Нижче подамо коротку характеристику набору даних ACDC.

– Пацієнти: 100 осіб із різними кардіологічними станами (норма, інфаркт міокарда, дилатаційна кардіоміопатія, гіпертрофічна кардіоміопатія, аномалії правого шлуночка).

– Зображення: Послідовні зрізи МРТ, що охоплюють весь серцевий цикл.

– Анотації: Ручні сегментування LV, міокарда та RV разом з експертними анотаціями ROI та орієнтирів.

– Поділ даних: Навчальна вибірка (70 пацієнтів), валідація (10 пацієнтів), тестування (20 пацієнтів).

### Процедура навчання моделі.

Модель глибокого навчання за архітектурою U-Net навчено через мінімізацію комбінованої функції втрат:

$$L_{total} = L_{seg} + \lambda_{anatomy} L_{anatomy} + \lambda_{aux} L_{aux},$$

де  $L_{seg}$  – функція втрат сегментування,  $L_{aux}$  – функція втрат для додаткових задач,  $\lambda_{anatomy}$  і  $\lambda_{aux}$  – вагові коефіцієнти.

Вагові коефіцієнти  $\lambda$  ініціалізовано емпірично, з огляду на продуктивність за валідаційним набором. Безпосередньо для навчання моделі використано оптимізатор Adam із початковою швидкістю навчання (learning rate)  $1 \times 10^{-4}$ . Також до процесу навчання було застосовано раннє завершення навчання та метод регуляризації для запобігання перенавчанню.

### Результати.

Результативність сегментування оцінено за коефіцієнтом Dice (DSC), індексом Жаккара (Jaccard), відстанню Гаусдорфа (HD) та середньою поверхневою відстанню (ASD).

В таблиці 1 подано результати обчислювальних експериментів. З таблиці 1 бачимо, що інтегрування експертних знань значно покращує продуктивність сегментування моделі U-Net. Використання механізмів уваги, керованих експертними анотаціями, дає змогу мережі фокусуватися на клінічно важливих областях, що підвищує якість подання ознак у цих регіонах. Анатомічні функції втрат забезпечують виконання реалістичних анатомічних співвідношень, що зменшує такі помилки, як некоректне маркування або перекриття структур.

Таблиця 1 – Порівняння результативності сегментації базової моделі U-Net та запропонованого підходу

Ділянка серця	Модель	DSC, % ↑	Jaccard, % ↑	HD, мм ↓	ASD, мм ↓
LV	Базовий U-Net	93.0 ± 1.6	87.0 ± 2.0	8.0 ± 2.2	1.2 ± 0.4
	Наш підхід	95.5 ± 1.2	91.5 ± 1.5	5.5 ± 1.8	0.8 ± 0.3
Міокард	Базовий U-Net	85.5 ± 2.5	76.0 ± 3.0	9.8 ± 3.1	1.5 ± 0.5
	Наш підхід	89.0 ± 2.0	81.0 ± 2.5	6.5 ± 2.3	1.0 ± 0.4
RV	Базовий U-Net	90.0 ± 2.1	82.0 ± 2.5	8.5 ± 2.5	1.3 ± 0.4
	Наш підхід	93.0 ± 1.7	86.5 ± 2.0	6.0 ± 2.0	0.9 ± 0.3

Порівняння підходів за усіма метриками сегментування ілюстровано на рисунку 1.

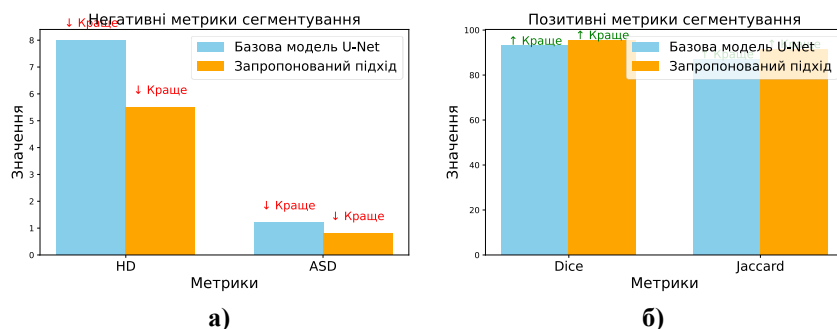


Рисунок 1 – Порівняння результатів сегментування моделей: запропонований підхід демонструє покращення за метриками Dice та Jaccard (позитивний тренд) і зниження помилок за метриками HD та ASD (негативний тренд) у порівнянні з базовою моделлю U-Net.

Рисунок 1 демонструє, що запропонований підхід перевершує базову модель U-Net за всіма ключовими метриками сегментації. Зокрема, вищі значення DSC та Jaccard свідчать про покращену якість сегментації, що краще відповідає реальним анатомічним структурам. Водночас, знижені значення HD та ASD вказують на менші геометричні похибки у сегментованих структурах, що є важливим для забезпечення анатомічної достовірності. Отже, інтеграція експертних знань, механізмів уваги та анатомічних обмежень у функції втрат забезпечує покращення результатів сегментації, що вказує на значну перспективу цього підходу для медичного діагностування за аналізом медичних зображень.

Нижче також подамо кілька обмежень запропонованого підходу, зокрема:

1. Залежність від експертних анотацій. Отримання анотацій може бути тривалим процесом і не завжди доступним у практичних умовах.
2. Додавання механізмів уваги та анатомічних втрат збільшує обчислювальні витрати.
3. Отримані кількісні результати є перспективними за набором даних ACDC, однак потрібна додаткова валідація за іншими наборами даних зображень MPT серця.

## Висновки.

У цій роботі запропоновано вдосконалений підхід до інтегрування експертних знань у модель глибокого навчання типу U-Net для сегментування зображень МРТ серця. Використання механізмів уваги, що керовані експертними анотаціями, та впровадження анатомічних функцій втрат дало можливість покращити сегментування проти базової моделі U-Net. Отримані результати вказують на підвищення точності, анатомічної достовірності та клінічної релевантності сегментації. Подальша робота буде зосереджена на зменшенні залежності від експертних анотацій та розширенні застосування цього підходу до інших завдань сегментування в медичній візуалізації.

## Список посилань.

1. *Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?* / O. Bernard et al. *IEEE Transactions on Medical Imaging*. 2018. Vol. 37, no. 11. P. 2514–2525. URL: <https://doi.org/10.1109/tmi.2018.2837502>
2. *Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science. Cham, 2015. P. 234–241.* URL: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
3. *Radiuk P. Applying 3D U-Net architecture to the task of multi-organ segmentation in computed tomography. Applied Computer Systems. 2020. Vol. 25, no. 1. P. 43–50.* URL: <https://doi.org/10.2478/acss-2020-0005>
4. *ACPL: Anti-curriculum pseudo-labelling for semi-supervised medical image classification* / F. Liu et al. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) : Proceedings, New Orleans, LA, USA, 18–24 June 2022. New York, NY, USA, 2022. P. 20697–20706.* URL: <https://doi.org/10.1109/cvpr52688.2022.02004>
5. *Chaban O. Integrating diagnostic models: A revolutionary approach in AI-driven healthcare. AI-Driven Transformation: Mapping the Course for Future Business Landscapes : Monograph / Gen. edit. O. Prokopenko, M. Jarvis, Tallinn: Teadmus OÜ, 2024, P. 204–216.* URL: <https://conference.euas.eu/2023/wp-content/uploads/2024/03/Monograph2023.pdf>

6. A survey on incorporating domain knowledge into deep learning for medical image analysis / X. Xie et al. *Medical Image Analysis*. 2021. Vol. 69. P. 101985. URL: <https://doi.org/10.1016/j.media.2021.101985>
7. Chaban O., Manziuk E. Enhancing medical NLI with integrated domain knowledge and sentiment analysis. *Information Control Systems & Technologies (ICST-2024) : CEUR-Workshop Proceedings, Odesa, Ukraine, 23–25 September 2024* / ed. by A. Pakštas et al. Vol. 3790. CEUR-WS.org, Aachen, 2024. P. 262–272. URL: <https://ceur-ws.org/Vol-3790/paper23.pdf>
8. Brauwers G., Frasincar F. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*. 2021. Vol. 35, no. 4. P. 3279–3298. URL: <https://doi.org/10.1109/tkde.2021.3126456>
9. High-level prior-based loss functions for medical image segmentation: a survey / R. El Jurdi et al. *Computer Vision and Image Understanding*. 2021. Vol. 210. P. 103248. URL: <https://doi.org/10.1016/j.cviu.2021.103248>
10. Чабан О.Р., Манзюк Е.А. Метод інтегрування доменних знань у багатостратегічну класифікацію медичних зображень. *Вісник Хмельницького національного університету. Серія «Технічні науки»*. 2024. Т. 337, № 3(2). С. 231–236. URL: <https://doi.org/10.31891/2307-5732-2024-337-3-34>

**Anders T. Sandnes, Bjarne Grimstad, Odd Kolbjørnsen** (Norwegian University of Science and Technology, Trondheim, Norway)

## MULTI-TASK LEARNING WITH LEARNED CONTEXTUAL INPUTS

**Abstract:** *This paper investigates learned-context neural networks, presenting a multi-task learning framework that utilizes a fully shared neural network structure augmented with an input vector containing trainable task-specific parameters. The architecture is notable for its efficient task adaptation mechanism, which leverages a low-dimensional task parameter space. Theoretically, it is demonstrated that a single scalar task parameter can suffice for universal approximation across all tasks—an advantage not necessarily shared by more conventional architectures. Empirical results reveal that, for similar tasks, the dimensionality of the task parameter correlates with task complexity, yet a compact parameter space remains feasible in most scenarios. Moreover, the task parameter space is characterized by stable and predictable behavior, facilitating workflows such as model updates with incoming data or learning new tasks while keeping shared parameters fixed. The architecture also exhibits resilience in handling datasets where individual tasks are represented by limited data points. Performance comparisons across ten datasets highlight its competitiveness against similar neural network architectures.*

**Keywords:** *learned-context neural networks, multi-task learning, task adaptation, universal approximation, low-dimensional parameter space, task-specific parameters, model updating, data efficiency.*

Our research focuses on addressing non-linear regression problems, particularly those involving multiple related tasks characterized by complex input-output dynamics and limited data availability. Such challenges are common in engineering and industrial systems, where tasks often share structural similarities. Examples include scenarios with repeated instances, like turbines in a wind farm, or batch processes, such as tracking biomass growth in aquaculture or agricultural fields. These tasks typically involve sparse data but are inherently similar by nature. To tackle these problems, an adaptable and highly flexible architecture is essential, enabling efficient task-specific adjustments even with minimal data. Moreover, the system must support practical operations necessary for maintaining machine-learning models, such as handling time-dependent variations or integrating new tasks that emerge over time. For instance, this might involve model updates to adapt to changing conditions or the incorporation of previously unobserved tasks.

We investigate neural networks that incorporate learned contexts. This architecture features two key components: a feedforward neural network with shared parameters across tasks and a set of task-specific parameter vectors. These task parameter vectors are additional inputs to the network, influencing its computations. Termed "learned contexts," they are optimized during training alongside the shared network.

Learned contexts offer an effective mechanism for task adaptation, achieving high adaptability with minimal task-specific parameters. Furthermore, they enable the discovery of a structured task parameter space—one that encapsulates the continuous latent characteristics of tasks rather than merely encoding task-specific information. Such a well-structured task parameter space is advantageous as it allows for training the shared network only once, shifting focus to task-specific parameters during daily operations. This approach proves particularly beneficial in scenarios where re-training the entire model is labor-intensive, computationally costly, involves frequent new data, or is constrained by limited access to the shared model [1,2].

Theoretical and experimental findings suggest that the learned-context neural network architecture is particularly well-suited for multi-task problems involving similar tasks or tasks with limited data. Its capability to represent tasks using low-dimensional and smooth parameter spaces makes it advantageous in such scenarios.

From a theoretical perspective, scalar task parameters are sufficient for a learned-context neural network to universally approximate all tasks, as demonstrated in Section 4. The use of contextual inputs enables meaningful task adaptations, even in compact network configurations. While the adaptability of the network improves with its size, this flexibility introduces a risk of

overfitting to specific tasks, which could undermine the benefits of multi-task learning. Therefore, careful tuning of hyperparameters is essential.

Experimentally, the optimal number of task parameters depends on the specific problem. However, the architecture generally supports extensive task adaptation using only a few parameters (Section 5.6). When training multiple tasks simultaneously, increasing the dimensionality of the task parameters often enhances performance (Section 5.7). Conversely, a smaller parameter space might be preferable when applying the shared model to new tasks. Practitioners must balance parameter dimensionality based on domain knowledge and the model's intended use. The architecture's ability to learn task parameters that capture latent task properties (Section 5.7, Appendix D) facilitates efficient updates and maintenance of the model in practice.

On full datasets, learned-context neural networks performed comparably to benchmark architectures (Section 5.4). On reduced datasets, their performance declined less sharply, indicating superior robustness (Section 5.5). All tested networks utilized standard fully connected layers, focusing on task adaptation rather than maximizing performance on specific datasets. Nevertheless, the learned-context approach could potentially be extended to other architectural designs.

The training process for learned-context networks has been robust, similar to that of other architectures (Section 5.4). The theoretical construction underpinning Theorem 2 suggests initializing task parameters to zero, as this avoids "task-encoding" local minima, which are more likely with random initialization. Zero initialization promotes similar tasks to follow aligned training trajectories, encouraging grouping of related tasks and fostering a well-structured parameter space. This reduces the risk of redundant parameter regions representing the same phenomena.

The study explores a fascinating architecture in the field of multi-task learning known as learned-context neural networks. This architecture is designed to address common challenges in machine learning, such as dealing with tasks that have limited data or require frequent updates due to dynamic conditions. By introducing trainable, task-specific parameters into a shared neural network, the approach achieves a balance between flexibility and efficiency, making it especially valuable in real-world scenarios where data may be sparse or tasks are interdependent. At its core, the learned-context neural network consists of a fully shared neural network augmented by trainable task parameters. These parameters act as contextual inputs, enabling the network to adapt to each task without requiring significant modifications to its structure. The ability to use low-dimensional task parameters ensures that the architecture remains computationally efficient while still being capable of capturing the unique characteristics of each task.

From a theoretical perspective, the study demonstrates the remarkable adaptability of this architecture. It is proven that even a scalar task parameter is sufficient for the network to approximate any set of tasks universally. This is an impressive feat, highlighting how learned-context neural networks can balance simplicity and power. Moreover, the architecture's design ensures that the task parameter space is well-behaved, meaning it represents latent properties of the tasks in a meaningful and interpretable manner. This is particularly useful for applications that require frequent updates, as it eliminates the need to retrain the entire network. The research also delves deeply into empirical evaluations, comparing learned-context neural networks with two well-established architectures: context-sensitive networks and last-layer neural networks. Across ten datasets, ranging from synthetic examples to real-world applications such as school performance analysis and disease monitoring, the learned-context networks consistently demonstrated competitive performance. They excelled in scenarios where tasks had limited data or were inherently diverse, showing less performance degradation compared to the other architectures.

One of the most striking observations was how the network's performance was influenced by the dimensionality of task parameters. While increasing the number of parameters generally improved performance, the benefits plateaued beyond a certain point. Interestingly, the architecture showed robustness even when task parameters were kept minimal, a feature that enhances interpretability and simplifies model management. For example, in datasets like "Sine and Line," where tasks could be represented with a few nonlinear basis functions, the network performed equally well with a minimal number of task parameters.

The training process was another area of focus. The learned-context neural network was shown to have a stable and reliable training mechanism. The study recommended initializing task parameters to zero, which facilitated grouping similar tasks and promoted smoother learning dynamics. This was particularly effective in preventing overfitting and ensuring that the task parameter space captured meaningful variations across tasks.

Additionally, the study explored how the architecture responded to reduced training datasets. Even when the data was drastically reduced to 10% of its original size, the learned-context networks maintained competitive performance, outperforming other models in most cases. This robustness makes the architecture highly suitable for practical applications where data availability may be limited.

The research also highlighted the architecture's potential for visualizing and interpreting task parameters. For example, in datasets related to health metrics like obesity or height, the task parameters aligned closely with domain-specific properties, offering insights into the underlying trends and patterns.

What sets this architecture apart is its adaptability and versatility. Unlike traditional multi-task learning approaches, learned-context neural networks do not rely on pre-defined task encodings or rigid structures. Instead, they dynamically adjust to the needs of each task, making them suitable for a wide range of applications. The study emphasizes that while the architecture was evaluated using fully connected neural networks, its principles could be extended to other designs, such as convolutional or transformer-based networks, opening doors for further innovations.

In conclusion, the learned-context neural network emerges as a robust and practical solution for multi-task learning challenges. Its ability to adapt to diverse tasks with minimal parameters, combined with its theoretical soundness and empirical robustness, positions it as a promising tool in both academic research and real-world applications. The study also paves the way for future exploration, suggesting that this architecture could be further refined and extended to tackle even more complex and dynamic learning problems.

## REFERENCES

1. Lester, B., Al-Rfou, R., & Constant, N. (2021). *The power of scale for parameter- efficient prompt tuning*. In M.-F. Moens, X. Huang, L. Specia, & S. W.-t. Yih (Eds.), *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 3045-3059). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.
2. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). *Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing*. *ACM Computing Surveys*, 55(9).



**Andreas Rauscher, Johannes Kaiser, Manoj Devaraju, Christian Endisch** (*Technische Hochschule Ingolstadt, Institute of Innovative Mobility (IIMo), Research Team Electromobility and Learning Systems (ELS), Ingolstadt, Germany*)

## DEEP LEARNING FOR PARTIAL DISCHARGE DETECTION IN ELECTRICAL MACHINES

**Abstract:** *Testing for faults in the production line of automotive traction machines is crucial to guarantee their expected lifespan. Repetitive partial discharges (PDs), triggered by anomalies in the insulation system, are a major cause of premature failures in electrical machines, making reliable PD detection highly significant. This study introduces deep learning (DL) techniques to enhance the differentiation of PD signals from background noise, outperforming conventional amplitude-based PD detection methods used in production lines. A systematic procedure for data extraction and labeling is outlined to ensure accurate datasets from various PD measurements. Additionally, the datasets are enriched with low signal-to-noise ratio PD pulses through a specialized data augmentation technique. A comparative analysis is performed on 13 different types of neural networks, including convolutional, recurrent, and fully connected architectures, using diverse time-frequency representations of the input signals. For each of the 13 network configurations, hyperparameters related to input transformation, network structure, and optimization solver are tuned to ensure a fair and comprehensive evaluation. The study reveals that a two-dimensional convolutional neural network combined with a continuous wavelet transform achieves the highest accuracy, approximately 99.76%, on a test dataset consisting of PD signals from previously untested objects. All DL models examined in this study outperform the traditional threshold-based PD classification approach. Notably, even for PD events with amplitudes close to the noise level, the best-performing model achieves a detection rate of about 95%. However, without the proposed data augmentation method, the DL models fail to reliably distinguish small PD pulses from background noise.*

**Keywords:** *Fault testing, Partial discharges, Deep learning, Data augmentation, Convolutional neural networks, Time-frequency representations, Continuous wavelet transform, Signal-to-noise ratio.*

End-of-line testing for traction machines is a critical step in ensuring the safety and quality of electric vehicles during mass production. Even minor defects in the insulation system, such as small cavities, can arise during manufacturing. When these weak spots are exposed to high voltage, they may result in localized breakdowns of the insulation, known as partial discharges (PD). These events not only accelerate the aging of the insulation but also lead to premature failures in electrical machines. To mitigate these risks, comprehensive PD testing of stator windings is mandated for all traction machines.

In this testing, surge voltages mimic real-world operational stresses caused by fast-switching inverters. Sensors like ultra-high-frequency antennas and high-frequency current transformers are employed to detect electromagnetic emissions and currents produced by PD events. However, differentiating PD signals from background noise in industrial environments poses a challenge due to the low signal-to-noise ratio (SNR). Achieving reliable PD detection is essential, both to ensure the robustness of the tested stators and to minimize costly and unsustainable false rejections. Conventional commercial PD detection systems typically use a straightforward amplitude-threshold approach. However, more advanced techniques have emerged, leveraging either feature-based machine learning or deep learning (DL). Unlike feature-based methods that rely on manual extraction of key data characteristics, DL approaches autonomously learn to identify relevant features from raw or transformed input data. This eliminates the need for expert-driven preprocessing and allows for greater accuracy and adaptability. Despite the advancements, PD testing under the specific conditions of non-sinusoidal voltages—required for traction machines—remains underexplored. Existing methods often use data transformations such as phase-resolved patterns or frequency-domain representations, but these are not well-suited for the high-speed surge voltages encountered in this context. To address these challenges, this research introduces DL methodologies tailored for PD classification in traction machines, emphasizing improved sensitivity and reduced false alarms. It also highlights innovations such as systematic data labeling and augmentation techniques to enhance the accuracy of small PD detection.

Neural Network Frameworks and Time-Frequency Transformations

This research begins by exploring the foundational concepts underlying neural networks and their specific implementations for the detection and classification of partial discharges (PD). Neural networks like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Fully Connected Neural Networks (FCNNs) play a crucial role in pattern recognition tasks. Each network type is tailored for unique strengths in processing time-sequential or spatial data.

### 1. Convolutional Neural Networks (CNNs)

CNNs leverage a hierarchical structure where convolutional and pooling layers extract and condense features from input data. By simulating the convolutional operation, feature maps are created, allowing CNNs to detect spatial patterns effectively. The combination of layers enhances the network's ability to discern increasingly complex patterns, which is vital for classifying PD signals amidst noise.

### 2. Recurrent Neural Networks (RNNs)

Designed for sequence data, RNNs capture temporal dependencies. Variants like Long Short-Term Memory (LSTM) networks address the challenges of vanishing gradients through gated mechanisms that regulate the flow of information and gradients. Bidirectional LSTMs (BiLSTMs) extend this by processing sequences in both forward and reverse directions, further enriching the contextual understanding of data. Another variant, Gated Recurrent Units (GRUs), simplifies the structure by reducing the number of gates, making them computationally efficient while retaining performance.

### 3. Time-Frequency Transformations

Time-frequency transformations, such as Short-Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), and Stationary Wavelet Transform (SWT), are employed to analyze signals in both time and frequency domains. These methods convert raw signals into representations that highlight key characteristics of PD events:

STFT partitions a signal into overlapping segments to analyze localized frequency changes.

CWT decomposes a signal across scales using wavelet functions, preserving fine temporal details.

SWT, unlike discrete wavelet transforms, avoids downsampling, ensuring consistency in temporal alignment.

These transformations enable neural networks to process signal data in a manner that enhances detection accuracy, especially for subtle PD events obscured by noise.

The study employs a sophisticated setup to test partial discharges (PDs) in automotive traction machines. The testing focuses on three-phase hairpin stators, which are critical components in these machines. Twelve stators were selected for the tests, with eight impregnated and four unimpregnated, categorized based on their insulation thickness and power classification. The distinction in insulation levels and power ratings allowed for a comprehensive analysis under varied operational conditions.

The experimental arrangement is meticulously designed. A surge generator (ST3810) equipped with a 10 nF capacitor serves as the high-voltage source. This generator simulates real-life stress by applying voltage surges to different phases of the stator windings. To capture the PD signals, ultra-high frequency (UHF) broadband antennas, including a horn antenna and a logarithmic-periodic antenna, are positioned strategically at both ends of the stator. The antennas are fine-tuned with analog high-pass filters to isolate the PD signals from distortions caused by the high-voltage pulses. Signal acquisition is performed using a state-of-the-art 12-bit digital oscilloscope operating at a sampling frequency of 10 GHz. This ensures high-resolution data collection for each surge event. Additionally, a high-voltage probe monitors the terminal voltage of the stators, adding another layer of precision to the testing procedure.

To ensure accuracy, the testing process adheres to the standards outlined in DIN IEC/TS 61934. This involves progressively increasing the peak voltage and meticulously recording the resulting antenna signals. The data, containing potential PD events, is subsequently processed in MATLAB for detailed analysis. This rigorous setup not only enhances the reliability of PD detection but also provides a robust foundation for generating datasets essential for further research in partial discharge classification. By addressing the challenges posed by diverse insulation conditions and

noise interferences, the setup ensures that the study's findings are both comprehensive and applicable to real-world scenarios. Creating a robust training dataset is critical for effectively applying neural networks (NNs) in the classification of partial discharges (PDs). However, due to the stochastic nature and random timing of PD events, reliably labeling such signals is a challenging task. The following methodology systematically addresses this issue by preprocessing raw data, isolating noise and PD events, augmenting data for smaller PD pulses, and optionally transforming the time-domain data into time-frequency representations.

### **Data Preprocessing**

Signals captured by antennas undergo preprocessing to prepare them for reliable analysis. This involves filtering out signals affected by clipping or saturation, as these may exceed the dynamic range of the data acquisition system. A digital high-pass Chebyshev IIR filter with a passband frequency of 200 MHz is applied to suppress distortions caused by high-voltage (HV) surges. Following this, signals are normalized relative to their noise levels. Noise levels are determined using data from the pre-HV surge interval, which excludes potential PD events. This normalization ensures a consistent thresholding process for subsequent PD extraction.

### **Noise Extraction**

The interval preceding the HV surge serves as a noise-only segment, providing a clean dataset for background noise analysis. By dividing this segment into smaller snippets of a specified length (e.g., 400 samples), a representative noise dataset is created. These snippets, devoid of PD events, serve as benchmarks to distinguish background noise from actual PD pulses.

### **PD Extraction**

To reliably extract PD events from the data, only signals surpassing a specific signal-to-noise ratio (SNR) threshold (e.g.,  $\text{SNR} > 4$ ) are labeled as PD. A window of data surrounding the identified peak is extracted and labeled. This process is repeated with random shifts in window placement to ensure robustness in model training, enabling the neural network to identify PD pulses even when they are not perfectly centered in the analysis window.

### **Dataset Augmentation**

The NN's ability to detect low-SNR PD events is enhanced through dataset augmentation. PD pulses are scaled down to various target SNR levels and overlaid onto noise snippets. This augmentation ensures the inclusion of weak PD signals in the training set, making the NN more adept at distinguishing subtle PD events from noise. By varying target SNRs across a range (e.g., 0.9 to 4), the augmented dataset captures a wide spectrum of PD characteristics.

### **Dataset Overview**

The datasets are carefully balanced to avoid class bias by randomly discarding excess samples from overrepresented categories. Separate datasets are designated for training, validation, and testing, with some exclusively reserved for evaluating generalization on unseen data. Additionally, data diversity is increased by incorporating signals from different antennas, stator configurations, and HV curve forms.

### **Data Transformation**

To explore alternative input formats, time-frequency transformations are applied to the datasets. Techniques such as Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT), and Stationary Wavelet Transform (SWT) are used. These methods provide additional insights by representing PD events in both time and frequency domains. The transformed data is then used as input for further analysis, enabling comparisons of classification performance across different representation methods.

This systematic process of data generation and transformation ensures that the NN models are trained on a comprehensive and representative dataset, optimizing their ability to accurately detect and classify PD events.

In this research, the process of network and hyperparameter tuning plays a pivotal role in optimizing the classification performance of neural networks for partial discharge (PD) detection. Initially, specific parameters such as batch size and solver type were assessed to identify their impact on training outcomes. This initial optimization revealed that different solvers, such as stochastic

gradient descent with momentum (SGDM) and adaptive moment estimation (ADAM), influenced performance based on the network type. For example, SGDM demonstrated superior results for certain convolutional neural networks (CNNs), whereas ADAM performed better for recurrent neural networks (RNNs) like LSTM and GRU. The next phase involved refining input transformations and network architectures. This step sought to balance computational efficiency with model complexity. Convolutional layers, known for their localized feature extraction capabilities, were configured with varying kernel sizes and layer depths. These configurations allowed the researchers to evaluate the trade-offs between network depth, the number of feature maps, and overall model accuracy. For recurrent networks, the number of layers and hidden units was adjusted to identify trends in performance improvements without overfitting.

A comprehensive comparison of 13 network configurations combined with input transforms such as Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), and Continuous Wavelet Transform (CWT) was conducted. Among these, the two-dimensional CNNs with CWT input consistently achieved the highest classification accuracy. The researchers optimized hyperparameters, including the learning rate, using Bayesian optimization to ensure robust performance across different datasets. The findings also highlighted the importance of preprocessing steps such as logarithmic transformations for specific input types like CWT. Additionally, applying data augmentation techniques significantly enhanced the network's ability to detect low-amplitude PD pulses, which are often obscured by background noise. Ultimately, the research demonstrated that carefully tuned CNN architectures could achieve detection accuracies exceeding 99%, even for challenging datasets. This marks a substantial improvement over traditional threshold-based PD detection methods, showcasing the potential of deep learning for improving reliability and precision in electrical machine diagnostics.

The study culminates in an extensive comparison of various deep learning models, designed to classify partial discharge (PD) signals with remarkable precision. Leveraging optimized hyperparameters, the researchers trained these models on a carefully curated dataset (Tr1) and validated their performance on unseen data (Te2). The overarching goal was clear: to push the boundaries of PD detection, surpassing traditional threshold-based methods in both accuracy and robustness. As the models were put to the test, the 2D Convolutional Neural Networks (2DCNNs) emerged as clear frontrunners. When paired with Continuous Wavelet Transform (CWT) or Short-Time Fourier Transform (STFT) input data, these networks achieved an impressive classification accuracy of nearly 99.75%. This level of precision was consistent across training, validation, and generalization datasets, proving the robustness and adaptability of the 2DCNNs. Interestingly, while other models like Gated Recurrent Units (GRU) and Bidirectional Long Short-Term Memory networks (BiLSTM) demonstrated strong performance, they slightly lagged behind the 2DCNNs. Notably, 1D Convolutional Neural Networks (1DCNNs) also performed admirably, particularly when working directly with raw data or utilizing Stationary Wavelet Transform (SWT) input. However, when FFT-transformed data was used, the accuracy dropped slightly, highlighting the importance of input transformation in maximizing model potential. The findings went beyond mere accuracy. A deep dive into the Receiver Operating Characteristics (ROC) of the best-performing models revealed exceptional true positive rates (TPR) above 99.5%, coupled with false positive rates (FPR) as low as 0.14%. These metrics underscore the reliability of the deep learning approaches in distinguishing PD events from noise—a critical factor in industrial settings where safety and cost-efficiency are paramount.

What's even more fascinating is how these advanced models fared against conventional threshold-based methods. While traditional techniques struggled with low signal-to-noise ratio (SNR) scenarios, the 2DCNNs maintained a detection rate exceeding 94% even for PD signals barely distinguishable from background noise. In contrast, threshold-based classifiers were prone to higher false rejection rates or lower detection accuracy, depending on the chosen sensitivity threshold. This study doesn't just highlight the superiority of deep learning over traditional methods—it paves the way for practical applications in industrial environments. By addressing challenges like noise interference and signal variability, these models hold the promise of enhancing production line

efficiency while ensuring the reliability of automotive traction machines. The 2DCNNs, in particular, stand out as a beacon of innovation, combining computational efficiency with unparalleled accuracy.

In the context of large-scale production of electrical machines, the importance of partial discharge (PD) testing has been emphasized as a critical step for ensuring the expected lifespan of automotive components. Traditionally, PD detection methods in production lines rely on amplitude threshold-based decisions as specified by DIN IEC/TS 61934. However, these methods struggle to distinguish PD signals with low signal-to-noise ratios (SNR) from background noise, leading to potential inaccuracies in classification.

This research offers a comprehensive solution by exploring 13 neural network (NN) architectures, including fully connected, convolutional, and recurrent networks, to determine the most effective deep learning (DL) approach for PD classification. These networks are enhanced by employing optional time-frequency transformations of input signals to improve their ability to differentiate between PD and noise. The study introduces a systematic methodology for data extraction, labeling, and augmentation, which significantly contributes to the quality and robustness of the training datasets. This process includes the generation of small PD pulses to ensure the models can accurately detect even subtle PD signals. Additionally, a meticulous hyperparameter optimization process was conducted, addressing factors such as batch size, solver selection, and learning rate, to refine network performance. The evaluation, carried out on datasets from previously unused test objects, revealed that two-dimensional convolutional neural networks (2DCNNs) utilizing short-time Fourier transforms (STFT) or continuous wavelet transforms (CWT) demonstrated superior accuracy, achieving approximately 99.75%. The 2DCNNs incorporate four convolutional layers and two fully connected layers in their classification head, trained with a stochastic gradient descent with momentum (SGDM) solver and a batch size of 16. Similarly, models like GRU and BiLSTM with stationary wavelet transform (SWT) inputs, along with 1DCNNs using raw data, performed almost as well, with accuracy differences of less than 0.1% compared to the 2DCNNs. The study also underscores the efficacy of the proposed data augmentation technique, which enables NNs to recognize small PD pulses more effectively. For instance, detection rates for signals with SNRs between 0.9 and 1 increased from 27% to 95% when data augmentation was applied. Furthermore, analysis through ROC curves demonstrated that all the investigated DL models surpassed traditional amplitude-based PD detection by achieving higher true positive rates (TPR) and lower false positive rates (FPR). This improvement directly contributes to enhanced product reliability and reduced false rejections in manufacturing.

Ultimately, the research highlights the potential of DL models to generalize across diverse PD signal types and noise conditions, paving the way for their integration into real-time production line applications. Despite the computational demands and the interpretability challenges associated with DL algorithms, the proposed approach marks a significant advancement in PD testing methodologies. The study concludes by suggesting further exploration into lightweight and advanced neural network architectures to optimize computational efficiency and expand application scopes.

**Antonello Cammarano, Vincenzo Varriale, Francesca Michelino, Mauro Caputo** (*University of Salerno, Fisciano, (SA), Italy*)

## **DISCOVERING TECHNOLOGICAL OPPORTUNITIES WITH NEURAL NETWORKS AND LITERATURE ANALYSIS.**

**Abstract:** *This study introduces a novel approach to identifying technological opportunities for advanced technologies by analyzing information extracted from academic literature. These opportunities are framed as potential business applications of technologies in novel, untested domains, such as industries or business processes that have yet to be explored. Unlike traditional methods that rely on alternative datasets, this approach emphasizes the value of leveraging emerging practices documented in scientific publications. The proposed methodology translates business cases described in research articles into a structured triad: technology, industry, and process, and evaluates their impact on business performance. From an extensive review of 33,285 papers, 14,739 distinct triads were identified. Using this dataset, an artificial neural network was developed to predict the potential impact of yet-to-be-tested combinations of technology, industry, and process. The methodology was applied to 11 advanced technologies, including 3D printing, artificial intelligence, blockchain, high-performance computing, digital applications, geo-spatial technologies, immersive environments, the Internet of Things, open and crowd-based platforms, proximity technologies, and robotics. For each technology, a "technological opportunity map" was created to pinpoint optimal untapped areas for future applications. Furthermore, the method differentiates between combinations with high and low confidence in expected impact, enabling businesses to prioritize the most promising opportunities. The paper concludes with a discussion on the implications of this methodology for both industry practitioners and academic research.*

**Keywords:** *technological opportunities, advanced technologies, business applications, literature analysis, neural networks, technology-industry-process triad, opportunity mapping, emerging practices.*

### **Introduction.**

Technological opportunities represent the untapped potential for advancing industries, markets, and products. Exploring such opportunities is a strategic priority for companies seeking long-term growth. By identifying and analyzing these opportunities, businesses can position themselves at the forefront of innovation, ensuring they remain competitive in an ever-changing landscape. Historically, companies have relied on systematic methods to predict and leverage technological trends. These include qualitative techniques like expert consultations and scenario planning, as well as quantitative methods utilizing patent data. While patents offer valuable insights into technological progress, they have limitations, particularly in the early stages of technological development, when opportunities often remain undefined. Moreover, patents focus on applied outcomes, potentially overlooking the foundational scientific knowledge that frequently drives innovation.

In contrast, academic literature offers a rich repository of emerging practices and theoretical advancements. Scientific research often precedes technological application, providing a unique vantage point for identifying nascent opportunities. This study leverages the potential of academic publications to bridge the gap between scientific exploration and business application. By focusing on case studies documented in peer-reviewed journals, the research creates a structured framework for identifying and evaluating the intersection of technologies, industries, and processes. Recognizing the growing influence of artificial intelligence and big data, this methodology integrates machine learning techniques to predict the impact of untested combinations. Through a detailed analysis of existing triads and their business implications, the study enables companies to make informed decisions about where to focus their efforts and investments. This narrative not only highlights the importance of embracing innovative approaches to technology forecasting but also underscores the transformative potential of using literature-based datasets. By aligning academic insights with practical applications, the study provides a roadmap for navigating the complex and dynamic landscape of technological advancement.

Understanding and harnessing technological opportunities is essential for fostering innovation. The process of identifying such opportunities involves analyzing how technological

advancements can reshape industries, businesses, and markets. However, this endeavor is often hindered by a lack of comprehensive information, especially during the early stages of technological development. When technologies are in their infancy, limited data restricts the ability to predict growth trajectories, apply curve-fitting techniques, or use stochastic models to anticipate future trends. This gap creates a significant divide between scientific exploration and the practicalities of commercialization and production.

Emerging technologies often face uncertainty regarding their developmental paths and potential commercial applications. Researchers, therefore, prioritize uncovering the most promising technological areas, guiding investments in specific companies, industries, or markets. Importantly, this exploration isn't limited to new technologies. Established technologies, even those beyond their early development stages, can also hold untapped potential. For this reason, the methodology described in this study was applied to 11 advanced technologies across varying phases of evolution. The analysis of technological opportunities benefits greatly from strategic intelligence and forward-looking assessment tools. These instruments enable decision-makers to chart pathways and evaluate success potentials. By anticipating emerging technologies and products, such assessments not only highlight innovations yet to be developed but also reveal new markets for existing technologies. This approach, broadly termed as "technological opportunity analysis," incorporates related concepts such as convergence analysis, vacant technology forecasting, and identification of emerging product or service opportunities. A key aspect of convergence analysis is its focus on the intersection of existing technological fields, where new inventions often arise. This dynamic evolution of technologies emphasizes the transformative potential of overlap between distinct domains, offering avenues for innovative applications in industries previously unlinked. Similarly, vacant technology analysis uses tools like maps or matrices to visually identify gaps in current technological implementations. These visualizations are instrumental in predicting potential connections between unassociated fields and directing efforts toward unexplored opportunities. Emerging and promising technologies, characterized by novelty, rapid growth, and transformational potential, are vital to this discourse. Such innovations often possess untapped market potential and can redefine industries. Accurately forecasting their effects before widespread adoption is critical, as uninformed investments carry significant risks. Researchers rely on this predictive capability to align decisions with the anticipated trajectories of technological progress. In parallel, identifying opportunities for products or services that align with technological advancements plays a foundational role in innovation. Such efforts guide the early stages of product or service development, optimizing business processes or extending product ranges to meet emerging market demands. Technological opportunity analysis can adopt either qualitative or quantitative methodologies. Qualitative methods, such as Delphi surveys or scenario planning, leverage expert insights but often suffer from subjectivity, time constraints, and limited scalability to specific industries or markets. Conversely, quantitative approaches rely on structured data from patents and scientific literature. Advances in artificial intelligence, machine learning, and natural language processing have enabled the analysis of large datasets, uncovering patterns and correlations that human experts might overlook. Patents have traditionally been the primary resource for such analysis, offering structured and standardized data. However, they have limitations, such as delayed availability, reliance on historical trends, and a narrow focus on existing innovations. In contrast, academic papers provide richer contextual insights, especially regarding new technologies in nascent stages. These publications often showcase the theoretical and practical implications of innovations, offering a broader perspective than patents. The strengths of academic literature lie in its coverage of diverse industries, detailed case studies, and emphasis on novel applications. By manually analyzing such papers, researchers can extract valuable insights into emerging practices, enabling a more nuanced exploration of technological opportunities. Although time-intensive, this approach mitigates the shortcomings of automated techniques, which often overlook deeper content within the body of scholarly works.

The methodology is organized into three main phases, each building upon the other to systematically uncover technological opportunities using cutting-edge technologies and academic research as a foundation.

In the **first phase**, researchers construct a dataset of "emerging practices" by conducting a thorough manual analysis of academic papers. These practices are categorized into triads consisting of technology, industry, and process, along with the expected impact on business performance. The meticulous analysis focuses on extracting detailed examples of how technologies have been implemented in specific contexts and the associated benefits. The aim here is to capture comprehensive insights into innovative applications from academic literature, ensuring that the dataset is rich with nuanced data.

The **second phase** utilizes this dataset to train an artificial neural network. By leveraging the triads as input, the model identifies patterns and relationships between technology, industry, and process, predicting their potential impact. The training process incorporates rigorous validation steps to ensure the model's reliability and accuracy. The network's task is to extrapolate the potential effects of combinations that have not yet been observed, essentially forecasting new opportunities for technological applications.

In the **final phase**, the trained neural network is deployed to assess untested combinations of technology, industry, and process. Each combination is analyzed for its predicted impact, categorized as either market-oriented (e.g., revenue growth or customer satisfaction) or organizational (e.g., cost savings or efficiency improvements). To provide actionable insights, the methodology introduces a confidence threshold for predictions, highlighting the most reliable opportunities while flagging those with higher uncertainty. Results are visualized through "opportunity maps," which graphically represent promising areas for future exploration, offering a clear and practical guide for businesses and researchers alike. The methodology is robustly tested across 11 advanced technologies, ranging from artificial intelligence and blockchain to 3D printing and robotics. By focusing on emerging practices and employing artificial neural networks, the approach bridges the gap between theoretical exploration and practical application, making it an invaluable tool for identifying high-impact technological opportunities.

Out of these, only 8,406 papers were deemed useful for the analysis, as the remainder either lacked relevant data on emerging practices or were inaccessible online. These selected papers facilitated the identification of 14,739 practices, encapsulated as combinations of technology, industry, and business processes. Interestingly, certain practices required the incorporation of multiple combinations to accurately capture their application, as they spanned various industries or processes. A diverse range of journals contributed to this rich dataset, including prominent publications like *Annals of Operations Research*, *Computers and Operations Research*, and *European Journal of Operational Research*. These sources demonstrated substantial variation in their focus on emerging practices, revealing trends in academic interest and technological applications. The artificial neural network, trained on the dataset, provided key insights into these combinations. Utilizing categorical variables such as technologies, industries, and processes, the network effectively predicted the potential market and organizational impacts of untested combinations. The neural network's architecture featured a single hidden layer optimized through SPSS, ensuring robust predictions tailored to real-world applications.

Across the evaluated technologies, 71.35% of potential combinations remain unexplored, signaling a significant opportunity for innovation. These vacant combinations, mapped through technology-opportunity frameworks, highlight promising areas where businesses can pioneer new applications. For instance, technologies like 3D printing exhibited higher shares of vacant opportunities, suggesting untapped potential in industries yet to adopt these innovations. Furthermore, the study underscored the pivotal role of processes in shaping outcomes. Sensitivity analysis revealed that process variables had the highest normalized importance, followed by industry and technology. This finding highlights the critical influence of how technologies are integrated into operational contexts to drive organizational and market performance. The methodology employed not only enhances our understanding of emerging practices but also provides actionable insights. By focusing on combinations with confident predictions of success, businesses can strategically prioritize areas with the greatest potential impact. This structured approach ensures that investments in new technologies are both informed and targeted, fostering a path toward transformative innovation.



The conclusion section of the paper outlines the innovative methodology proposed for uncovering technological opportunities using insights derived from scientific literature and artificial neural networks. This approach demonstrates significant predictive power, as validated through testing across 11 cutting-edge technologies. However, the methodology has its limitations, primarily tied to the nature of the dataset and the reliance on manual content analysis. This time-intensive process stems from the substantial volume of scientific literature published annually and the varying relevance of these works to specific technological applications. Furthermore, the absence of certain applications in the database might lead to an overestimation of potential combinations. As new technologies emerge, integrating them into the existing classification and retraining models from scratch become necessary, posing additional challenges. The study emphasizes that scientific literature serves as a valuable repository for identifying emerging practices, allowing for the development of tools that uncover technological opportunities. Future research should focus on exploring individual technologies in greater depth, as listed in the supplementary tables, to refine the methodology further. Continuous monitoring of scientific literature is essential to incorporate newly emerging technologies and assess their commercial viability and developmental progress. To mitigate the challenges of manual analysis, efforts are underway to integrate advanced natural language processing tools that can pre-filter relevant abstracts and prioritize the most promising papers. This automation aims to significantly reduce the time required for analysis. Moreover, with advancements in artificial intelligence, it will soon be feasible to automate the extraction of emerging practices directly from scientific texts, streamlining the process and enabling quicker identification of technological opportunities.

In summary, while the methodology outlined is robust and offers substantial insights, its application can be further enhanced through ongoing technological advancements and methodological refinements. This iterative process ensures the approach remains relevant and capable of addressing the dynamic landscape of technological innovation.

**Carmen Bisogni, Lucia Cimmino, Michele Nappi, Toni Pannese, Chiara Pero** (*University of Salerno, Department of Computer Science & Department of Management & Innovation Systems, Fisciano, Salerno, Italy*)

## **GAIT-BASED EMOTION RECOGNITION WITH PRIVACY PRESERVATION**

**Abstract:** *Recognition of emotions based on gait has attracted growing attention due to its diverse applications in areas such as healthcare, social interaction analysis, surveillance systems, and smart technologies. Unlike other biometric traits, gait offers distinct advantages, including the ability to perform remote identification and maintain reliable performance even in unstructured environments. Additionally, analyzing gait patterns can provide meaningful insights into an individual's emotional condition. This study introduces the "Walk-as-you-Feel" (WayF) framework, an innovative method for emotion recognition from gait that prioritizes user privacy by avoiding reliance on facial information. To overcome challenges associated with limited and imbalanced datasets, the approach incorporates a tailored balancing technique designed for deep learning models. Advanced architectures like Adapted Inception-v3 and EfficientNet are utilized for feature extraction, while emotion classification is achieved through Gated Recurrent Units (GRUs) and a Transformer Encoder. The experimental evaluation highlights the effectiveness of WayF, achieving an average recognition accuracy of approximately 77% in its optimal configuration. Furthermore, when excluding neutral emotions, the system attains an impressive accuracy of 83.3%, surpassing many methods that incorporate facial features.*

**Keywords:** *gait, emotion recognition, biometric, privacy, deep learning, feature extraction, grus, transformers*

### **Introduction.**

Understanding human emotions through various recognition techniques has become an increasingly significant topic across numerous domains. Emotions play a pivotal role in enhancing human interaction, influencing decision-making, and shaping behavioral patterns. Fields such as healthcare, social communication, surveillance, and advanced technology applications have particularly benefitted from advancements in emotion recognition. Researchers have explored multiple biometric approaches to detect emotional states, including speech analysis, facial expression recognition, EEG signals, and gait patterns. Among these, gait analysis has emerged as a distinct and promising method for emotion recognition.

Gait, fundamentally defined as the pattern of an individual's movement during walking, transcends its basic role as a means of mobility. It serves as a unique identifier, capable of distinguishing individuals based on their specific walking styles. This biometric trait offers remarkable advantages over traditional methods, such as remote identification, which does not require physical proximity, and effectiveness under diverse, uncontrolled conditions. Gait recognition systems can operate efficiently even with low-resolution data and do not rely on the active cooperation of individuals. This makes gait-based identification a practical choice for real-world scenarios, where conventional techniques like facial recognition might fail due to obscured or masked features.

Beyond identification, gait analysis provides profound insights into emotional states, bridging the gap between physical behavior and affective conditions. Variations in gait patterns are known to correlate with different emotional states, making it possible to infer an individual's mood or psychological disposition through their walking behavior. This interplay between gait dynamics and emotions introduces innovative opportunities for emotion recognition technologies, particularly in privacy-sensitive applications.

Unlike facial recognition, which directly captures and interprets expressive features, gait analysis offers a less intrusive alternative, safeguarding personal privacy while delivering actionable insights. This unique aspect is particularly relevant in environments where individuals may intentionally mask their faces or where privacy concerns limit the use of facial-based recognition systems. By focusing solely on gait patterns, researchers aim to eliminate biases and limitations associated with face-dependent methods, paving the way for broader adoption of emotion recognition systems in surveillance, healthcare monitoring, and beyond.

Recent advancements in gait analysis have delved into optimizing feature extraction and improving resilience against intra-class variations. Despite these developments, the field faces challenges, particularly when excluding facial cues, which are traditionally strong indicators of emotional states. Addressing these challenges requires innovative approaches that not only enhance the accuracy of gait-based emotion recognition but also respect ethical considerations such as user privacy.

In this context, the "Walk-as-you-Feel" (WayF) framework introduces a groundbreaking solution. This novel approach leverages advanced deep learning architectures to analyze gait patterns while deliberately excluding facial information. By prioritizing privacy-preserving mechanisms, the WayF framework broadens the scope of biometric applications, offering a robust method to decode emotions solely from walking behavior. The research presents a multifaceted contribution, including the development of a dataset balancing procedure to address limitations posed by small and uneven data distributions, and the integration of cutting-edge feature extraction models like Inception-v3 and EfficientNet, combined with Gated Recurrent Units (GRUs) and Transformer Encoders for classification.

Through rigorous experimentation, the WayF framework demonstrates promising results, achieving competitive accuracy levels even when benchmarked against state-of-the-art methods that utilize facial cues. This work not only highlights the potential of gait as a standalone biometric for emotion recognition but also underscores the importance of privacy-centric design in modern AI systems.

### **Narrative Summary of Related Works**

Emotion recognition plays a crucial role in various fields, such as psychology, healthcare, communication, and technology, offering deeper insights into human experiences and improving engagement. Among the methodologies for emotion recognition, two dominant approaches have emerged: video-based analysis and skeleton-based analysis. Each provides unique advantages, with video-based approaches capturing a broad range of emotional cues, including facial expressions, while skeleton-based methods focus on key body movements, offering a more abstract yet privacy-preserving perspective.

#### **Video-Based Approaches**

Recent advancements in video-based emotion recognition have leveraged gait analysis alongside deep learning architectures. For instance, Bhattacharya et al. introduced a Bi-Modal Deep Neural Network (BMDNN) that combines domain-specific discriminative features with deep latent features, achieving state-of-the-art results on the Edinburgh Locomotion MoCap dataset. Similarly, Yin et al. developed a MultiScale Adaptive Graph Convolution Network (MSAGCN) capable of dynamically selecting spatial-temporal features, outperforming previous techniques in accuracy. Another notable contribution by Chai et al. involved the use of multi-head pseudo-nodes to address feature imbalance and enhance the global understanding of joint interactions. These approaches highlight the potential of video-based methods to decode complex emotional states by analyzing a range of visual cues.

#### **Skeleton-Based Approaches**

In the realm of skeleton-based emotion recognition, researchers have explored innovative strategies to extract emotional insights from body movements. Karg et al. demonstrated the feasibility of recognizing emotions through gait patterns by using dimensionality reduction techniques like Principal Component Analysis (PCA) and achieving high accuracy with person-specific data. Venture et al. extended this approach by analyzing joint angles derived from motion capture data, revealing promising results in classifying emotions.

Crenn et al. introduced a method that combines body posture analysis with motion features, employing a Support Vector Machine (SVM) classifier to distinguish emotional expressions effectively. Expanding on this, the team later developed a system that computes spectral differences between neutral and expressive motions, further refining the classification process. Similarly, Daoudi et al. proposed analyzing temporal dynamics of skeleton joint movements in 3D space, utilizing covariance matrices and Riemannian geometry for emotion classification.

These works collectively emphasize the richness of gait and skeletal data in decoding emotions, providing robust alternatives to traditional facial expression analysis. Each approach contributes to advancing the field, highlighting unique challenges and opportunities for future exploration in emotion recognition.

### **Methods**

The methodology adopted in this study focuses on leveraging advanced deep learning architectures to analyze gait patterns for emotion recognition, ensuring user privacy by excluding facial cues.

#### **Transformers-Encoder**

Initially introduced in the seminal paper “*Attention Is All You Need*” (Vaswani et al., 2017), Transformers have revolutionized sequence data analysis by overcoming limitations of Recurrent Neural Networks (RNNs). Their self-attention mechanism enables them to identify relationships across distant elements within a sequence, making them particularly well-suited for processing variable-length input data. For this study, only the encoder component of the Transformer architecture is utilized, as it efficiently extracts discriminative features for classification tasks. The use of Transformers in the “Walk-as-you-Feel” (WayF) framework exemplifies an innovative approach in computer vision by exploring their application beyond natural language processing.

#### **Convolutional Neural Networks**

Convolutional Neural Networks (CNNs) are employed for spatial feature extraction. Two architectures—EfficientNetV2-L and Inception-v3—are utilized in different pipeline configurations. EfficientNetV2-L, known for its superior design integrating progressive learning and advanced architectural modifications, excels in performance when paired with Transformers-Encoder. Inception-v3, which emphasizes efficient parameter usage and advanced factoring strategies, shows robust compatibility with RNN-GRU networks.

#### **Gated Recurrent Units (GRUs)**

GRUs, a variant of recurrent neural networks, are selected for their efficiency in handling sequential data while mitigating issues like vanishing gradients. With a simpler architecture compared to LSTMs, GRUs utilize Update and Reset gates to manage information flow and ensure robust temporal feature processing. This makes them a reliable choice for tasks requiring detailed sequential analysis, such as gait-based emotion classification.

#### **MediaPipe Pose Landmark Detection**

For landmark extraction, the MediaPipe library is employed to predict 33 body points from video frames. By excluding facial features, the framework emphasizes body motion, ensuring privacy preservation. The extracted body landmarks serve as inputs for the deep learning pipelines.

#### **Pipeline Design**

Two distinct pipelines are developed:

1. **EfficientNetV2-L + Transformers-Encoder:** This pipeline excels in capturing temporal features through the Transformer’s attention mechanism, particularly effective for neutral and sad emotions.
2. **Inception-v3 + RNN-GRU:** Designed to process sequential data, this configuration shows strong performance for happy and angry emotions.

Both pipelines involve pre-processing the video frames into feature vectors, which are subsequently classified by either the GRU or Transformer network. The selection of 240 frames per video ensures sufficient data for capturing intricate gait patterns.

#### **Experimental Protocol: Narrative Overview**

The experimental design aimed to evaluate the feasibility of emotion recognition using gait data while maintaining user privacy. The study utilized the E-Walk dataset, chosen specifically for its absence of facial information. This unique dataset ensured the focus remained solely on gait-based analysis without relying on facial features, a crucial aspect of preserving privacy.

#### **Dataset Characteristics and Preprocessing**

The E-Walk dataset includes 84 video clips featuring various walking styles. Each video was rated by multiple participants on a scale of 1 to 5 for emotions such as happiness, sadness, anger, and

neutrality. The scoring system averaged these ratings to classify each video by the dominant emotion. A critical preprocessing step involved extracting 240 frames from each video using the OpenCV library, resulting in a total of 20,160 frames. Subsequently, the MediaPipe Pose Landmark Detection library was employed to extract body landmarks, generating gait masks that represented body movements. This process excluded facial landmarks to ensure privacy and minimize the potential for misclassification caused by facial occlusion.

### **Data Balancing Strategies**

A major challenge addressed in this study was the imbalance inherent in the dataset, with certain emotions being overrepresented. To overcome this, the researchers employed both under-sampling and over-sampling techniques. Under-sampling reduced the dominance of overrepresented classes, while over-sampling replicated instances of underrepresented emotions. This balancing ensured that each emotional category contributed equally during training, enhancing the model's robustness.

### **Implementation of Pipelines**

Two distinct machine learning pipelines were developed for the analysis:

1. **CNN + Transformers-Encoder Pipeline:** This pipeline integrated EfficientNetV2-L as the convolutional neural network (CNN) for feature extraction, paired with Transformers-Encoder blocks for classification. The Transformer's auto-attention mechanism and internal masking provided nuanced temporal analysis of gait patterns.
2. **CNN + RNN-GRU Pipeline:** This second pipeline employed the Inception-v3 network for CNN-based feature extraction, followed by classification using Gated Recurrent Units (GRUs). The GRU architecture was chosen for its simplicity and effectiveness in handling sequential data.

Each pipeline processed the feature vectors generated by the CNN, enabling the downstream classification algorithms to discern subtle temporal correlations.

### **Experimental Findings**

The experiments were conducted under two conditions: balanced and unbalanced datasets. In balanced setups, data was divided into training and testing sets in an 80:20 ratio. This configuration allowed for a rigorous evaluation of the models while maintaining sufficient training samples. Notably, the CNN + Transformers-Encoder pipeline demonstrated superior performance for certain emotions, such as neutrality, while the CNN + RNN-GRU pipeline excelled in recognizing happiness. Together, these pipelines highlighted complementary strengths.

When the neutral emotion was excluded—due to its inherent ambiguity in psychological contexts—the best-performing configuration achieved an accuracy of 83.3%, underscoring the framework's potential.

### **Conclusions**

This research introduces the innovative "Walk-as-you-Feel" (WayF) framework, pioneering an approach for recognizing emotions through gait analysis, prioritizing privacy by excluding facial features. By leveraging body movement data, the method focuses on sequences derived from skeletal motion. To overcome the inherent challenges of small and unbalanced datasets often encountered in emotion recognition tasks, this study incorporates tailored balancing techniques optimized for deep learning systems. The architectural foundation of WayF is built on adapting state-of-the-art convolutional neural networks (CNNs) such as Inception-v3 and EfficientNet, which serve as feature extractors. These features are then processed by advanced classifiers like Gated Recurrent Units (GRU) and Transformers-Encoder models. Through rigorous experimentation, the WayF framework achieved competitive results, with an average accuracy of 77%, a figure that rivals existing methods which incorporate facial data. Significantly, when the "neutral" emotional state is omitted, the accuracy improves to 83.3%. These findings underscore the potential of gait-based emotion recognition, particularly in scenarios where facial cues are unavailable or intentionally excluded to ensure user privacy. The reliance on skeletal data expands the usability of this method in diverse applications, from healthcare to surveillance, while maintaining ethical considerations regarding personal data.

Future research directions include exploring hybrid datasets, such as the partial incorporation of synthetic data like E-Gait, which could provide broader benchmarks for model training. Such endeavors would necessitate modifications to ensure the dataset aligns with WayF's emphasis on non-facial analysis. Additionally, conducting ablation studies on the impact of individual emotions, informed by E-Walk's detailed emotion scoring, could provide deeper insights into the emotional nuances of gait. This shift from a purely classification approach to potentially integrating regression modeling could enhance the framework's predictive capabilities. Further expansion of the research could involve integrating multimodal data, such as audio or physiological signals, which may provide complementary insights into emotional states. Evaluating the system's real-time performance in dynamic environments will also be critical in determining its practical viability and robustness.

**Daniel H. Stolfi, Grégoire Danoy** (*Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg*)

## **EVOLUTIONARY SWARM FORMATION FOR AUTONOMOUS ROBOTS**

**Abstract:** *Autonomous robot swarms have emerged as a promising solution for various space and aerospace applications due to their adaptability, resilience, and ability to self-organize. These systems are particularly useful in tasks such as asteroid surveillance, convoy protection, and counter-drone operations, which require maintaining stable formations around a central target. However, the challenge lies in managing swarms with varying robot numbers and diverse initial conditions. In this research, we introduce an innovative method for the self-organization of autonomous robotic swarms, where individual robot movements are guided solely by relative positional data (range and bearing) derived from radio beacons. To optimize swarm behavior, we employ an evolutionary algorithm to determine optimal parameters, such as speed and attraction/repulsion forces, ensuring robust formation stability across different initial setups and failure scenarios. We tested this approach through realistic simulations across six scenarios, involving swarms of 3, 5, 10, 15, 20, and 30 robots. The results demonstrated the effectiveness of the optimized configurations, achieving stable circular formations in all 420 test scenarios. Additionally, we validated the method using physical experiments with E-Puck2 robots, confirming the swarm's ability to self-organize around a central target and maintain formation despite robot failures, achieving consistent success in all trials.*

**Keywords:** *autonomous robots, robot swarms, self-organization, evolutionary algorithms, swarm optimization, stable formations, aerospace applications, robot resilience*

### **Introduction.**

Autonomous robot swarms have emerged as a viable solution for space and aerospace applications due to their adaptability, robustness, and self-organizing abilities. These swarms are particularly useful in tasks like asteroid monitoring, convoy protection, and counter-drone operations, where stable formations around a central target are essential. However, managing varying swarm sizes and diverse initial conditions presents a significant challenge.

In this research, we propose a novel method for organizing autonomous robot swarms, where each robot's movement is determined solely by its relative position, measured through range and bearing from radio beacons. We employ an evolutionary algorithm to optimize the swarm's parameters—such as speed and attraction/repulsion forces—to ensure stable formations under various conditions, including different initial setups and failure rates. The method is tested using realistic simulations with swarms of varying sizes (3, 5, 10, 15, 20, and 30 robots), with the optimized configurations consistently achieving circular formations across multiple test scenarios. Additionally, real E-Puck2 robots were used for validation, demonstrating the swarm's ability to self-organize and maintain stability despite robot failures.

### **Optimization Approach and Methodology**

In this study, we present an innovative method for the self-organization of autonomous robot swarms, where the robots' movements are solely determined by their relative positions to each other. These positions are measured through range and bearing data, which is obtained using their respective radio beacons. This approach eliminates the need for central control or global positioning systems, allowing the robots to coordinate efficiently without requiring extensive infrastructure.

To enhance the swarm's ability to maintain robust formations under a variety of conditions, we introduce an optimization strategy based on an evolutionary algorithm. The goal is to determine the ideal parameters for the swarm, including the speed of movement and the forces of attraction and repulsion between robots. These parameters are crucial for ensuring that the swarm can adapt to different initial configurations, environmental conditions, and potential robot failures, while still achieving stable and predictable formations.

The evolutionary algorithm is designed to explore the search space and identify the best values for these parameters. By evaluating the fitness of different swarm configurations, the algorithm fine-tunes the system to achieve the desired formation in the most efficient way possible. This method

ensures that the swarm can operate effectively even in the presence of uncertainties or disturbances, such as changes in the number of robots, their initial positions, or unexpected failures.

In essence, our approach provides a scalable and adaptive solution for organizing autonomous robots into stable formations, while maintaining flexibility and resilience. The optimization process ensures that the swarm can function optimally under a wide range of conditions, making it a highly effective strategy for applications that require dynamic coordination, such as space exploration or security operations.

In this research, we introduce an innovative method for the self-organization of autonomous robot swarms, where the movement of each robot is determined solely by its relative position in the swarm. This relative position is derived from range and bearing information obtained through their respective radio beacons. The robots do not rely on an external control system or central coordination. Instead, their movements are based on local interactions with neighboring robots, allowing the swarm to collectively achieve the desired formation.

To further enhance the swarm's performance, we propose an optimization approach that uses an evolutionary algorithm to calculate the optimal parameters necessary for maintaining robust formations. These parameters include factors such as the robots' speed and the attraction/repulsion forces between them. The goal is to find the most effective configuration that ensures the stability of the swarm's formation, even under varying initial conditions and in the presence of potential failures. By optimizing these parameters, the swarm can adapt to different scenarios, maintaining its structure and performance despite challenges such as robot malfunction or changes in the swarm size.

This approach addresses the complexity of managing autonomous robot swarms, where factors like the number of robots, the starting conditions, and the potential for failure introduce significant challenges. The proposed method aims to overcome these difficulties by providing a solution that is both scalable and resilient. Through the optimization of swarm parameters, it is possible to create stable and adaptable formations, even in dynamic environments where the robots must rely on local information and interactions.

Unlike traditional methods that rely on centralized control, our approach allows each robot to adjust its movement based solely on its relative position, which is determined through range and bearing measurements obtained from radio beacons. The robots do not need to communicate with one another directly or rely on complex algorithms; instead, they use simple local information to coordinate their actions.

To ensure the formation remains stable and robust, we employ an evolutionary algorithm to optimize key swarm parameters. These include the speed of each robot as well as the attraction and repulsion forces between them. The goal is to maintain a stable, organized swarm that can adapt to varying initial conditions and account for potential robot failures. This is particularly important as the swarm must be resilient, even when some robots malfunction or become disconnected from the rest.

We tested this approach through a series of experiments that simulated six different scenarios, with varying numbers of robots ranging from three to thirty. In total, 420 scenarios were run, and the results were consistently successful. In every case, the swarm achieved the desired circular formation, demonstrating the effectiveness and stability of the approach.

Finally, to validate the method in a real-world environment, we used physical robots, specifically the E-Puck2 models, to conduct further experiments. These real-world tests confirmed that the swarm was capable of self-organizing around a central point of interest. Additionally, the swarm showed resilience to robot failure, as it continued to maintain its formation even when individual robots failed, further proving the robustness of our approach.

In this section, we present a novel approach for optimizing the behavior of autonomous robot swarms, focusing on their self-organization capabilities. The key innovation lies in how the robots move: rather than relying on external commands or predefined paths, each robot's movement is based purely on its relative position to other robots in the swarm. This positional information is derived from the range and bearing measurements, which are obtained through radio beacons. These relative measurements allow the robots to coordinate their actions without needing a central controller, mimicking natural swarming behaviors.



To fine-tune the swarm's movement dynamics, we apply an optimization technique grounded in evolutionary algorithms. This approach helps determine the ideal parameters for the swarm, such as the speed of the robots and the forces of attraction and repulsion that dictate their interactions. By optimizing these parameters, we aim to ensure the swarm can form and maintain stable and robust formations, even in the face of changing initial conditions and potential robot failures.

The effectiveness of this method was tested through a series of realistic simulations. These simulations featured various swarm sizes, ranging from three to thirty robots, across six distinct case studies. Each scenario aimed to challenge the swarm's ability to form stable circular formations under different conditions. The results were promising: the optimized configurations consistently enabled the swarm to achieve the desired formation in 420 test cases. The success of these simulations demonstrated that the proposed method is resilient and effective, capable of achieving robust performance even when faced with varied starting conditions.

Lastly, the research was extended to real-world testing using E-Puck2 robots, which were used to physically validate the swarm's ability to self-organize. In these experiments, the robots successfully formed stable circular formations around a central point of interest. Furthermore, the swarm demonstrated impressive resilience, maintaining its formation even when individual robots failed, confirming the robustness and reliability of the proposed self-organizing system.

## **Conclusion.**

To ensure that the swarm achieves robust formations, we employ an optimization strategy based on an evolutionary algorithm. This algorithm is used to calculate key parameters for the swarm, such as speed and the forces of attraction and repulsion between the robots. By optimizing these parameters, the swarm can maintain stable and efficient formations even when faced with varying initial conditions or failure rates.

We tested this method through simulations designed to reflect real-world scenarios. These simulations involved six different case studies, with the swarm size ranging from three to thirty robots. The results from the simulations were promising, with the optimized configurations consistently achieving stable circular formations across all 420 test scenarios. This success demonstrated the robustness of our approach under various conditions.

Finally, we validated our method through practical experiments using real E-Puck2 robots. The physical tests confirmed that the swarm could effectively self-organize around a central point of interest, even when individual robots experienced failures. In every trial, the swarm was able to maintain the desired circular formation, further supporting the reliability and resilience of the proposed method.

**Daniel Marek, Jakub Nalepa** (*Department of Algorithmics and Software, Silesian University of Technology, Gliwice, Poland; KP Labs, Gliwice, Poland*)

## DEEP LEARNING FOR EX-TRATERRESTRIAL ROCK SEGMENTATION

**Abstract:** *Bringing autonomy to edge devices is a crucial step in advancing space exploration. Among the many tasks that autonomous vehicles can perform, detecting and segmenting rocks in onboard images of extraterrestrial landscapes is pivotal for enabling safe navigation and avoiding collisions. To address this challenge, we propose an end-to-end pipeline designed to develop and validate resource-efficient machine learning techniques tailored to this task, offering significant flexibility. Deploying such models on edge devices involves numerous practical challenges, including achieving memory and computational efficiency and ensuring robustness under varying image quality conditions. These critical aspects are often neglected during the development of deep learning-based onboard systems. We demonstrate that these considerations should be integral to the deployment process. Our extensive experimental analysis across multiple benchmark datasets highlights the functional and non-functional capabilities of the models, both in full-precision and in compressed formats via quantization, the latter achieving similar segmentation accuracy while reducing model size by approximately 11 times. Furthermore, we illustrate that synthesized images can be effectively used to evaluate the robustness of deep learning models under conditions mimicking onboard acquisition scenarios, where degraded image quality can negatively impact the performance of models trained on clean, high-quality data.*

**Keywords:** *autonomy, edge devices, space exploration, rock segmentation, machine learning, robustness, quantization, image quality*

Exploration of space presents humanity with extraordinary challenges. The harshness of extraterrestrial environments—marked by cosmic radiation, extreme temperatures, and the absence of breathable air—necessitates the development of highly specialized vehicles like rovers. These rovers are designed not just to survive but to perform a variety of critical tasks, such as terrain analysis, collecting samples, testing mission equipment, and transmitting invaluable data back to Earth.

However, the vast distances of space introduce a unique constraint: the time delay in radio signal transmission makes real-time remote control impossible. As a result, these machines require a significant degree of autonomy. While current rovers depend partially on commands from Earth, transitioning to full autonomy promises to revolutionize their efficiency and effectiveness in fulfilling mission objectives. Autonomous navigation, hazard detection, and power management are among the essential capabilities that would enable rovers to adapt seamlessly to their environment.

Despite their potential, designing these systems for space exploration involves overcoming stringent technical challenges. The onboard computers must operate under severe hardware constraints, balancing limited processing power with optimal performance. To navigate these limitations, advanced machine learning techniques have become indispensable. Unlike traditional algorithms, machine learning models excel in capturing complex relationships within data, providing a robust foundation for autonomous decision-making under uncertain conditions.

This work focuses on an essential aspect of rover autonomy: the detection and segmentation of rocks from onboard images. This capability not only aids in collision avoidance but also facilitates path planning and other critical functions. Leveraging recent advances in deep learning, we present an innovative end-to-end pipeline tailored for resource-constrained edge devices. By addressing both the functional and non-functional aspects of deploying these models in extraterrestrial conditions, we aim to bridge the gap between theoretical advancements and practical implementation.

The integration of deep learning models into the hardware of rovers poses a series of unique challenges. These models must not only provide accurate segmentation of rocks but also operate efficiently within the severe computational and memory limitations of onboard devices. Achieving this balance requires innovative solutions that optimize the size and performance of the models without compromising their accuracy. To address these challenges, we introduce a systematic approach that combines state-of-the-art machine learning techniques with advanced resource optimization strategies.

A critical aspect of our approach involves tailoring the models to the specific conditions of extraterrestrial environments. Images captured by rover cameras are often affected by noise, varying lighting conditions, and other factors that degrade their quality. Traditional models trained on clean, high-resolution datasets may falter under such conditions. To ensure robustness, our pipeline incorporates data-level simulations that mimic these real-world acquisition scenarios. By introducing controlled noise, blur, and exposure variations during the training process, we enhance the ability of our models to generalize across diverse environmental conditions.

In addition to ensuring robustness, we focus on compactness through post-training quantization. This process reduces the size of the models significantly—often by an order of magnitude—without a noticeable drop in performance. The resulting models are not only suitable for deployment on memory-constrained devices but also exhibit faster inference times, which is crucial for real-time applications like navigation and hazard avoidance.

Our extensive experimental evaluation spans multiple benchmark datasets, including both synthetic and real-world images of extraterrestrial landscapes. These datasets capture the diversity of conditions that rovers might encounter, from the barren expanses of the Moon to the rugged terrains of Mars. By testing our models under a variety of scenarios, we demonstrate their adaptability, precision, and efficiency. The results confirm that our pipeline offers a viable solution for integrating deep learning-powered rock segmentation into the operational framework of space exploration vehicles.

To promote transparency and reproducibility, we have made our entire implementation publicly available. This allows other researchers to build upon our work, adapt our methods to different applications, and further advance the field of autonomous systems for space exploration. As we look to the future, we envision this pipeline serving as a foundational tool not only for rock segmentation but also for a range of other onboard image analysis tasks.

---

In the broader context of space exploration, the implications of deploying such compact and robust machine learning models are profound. As missions extend to more distant celestial bodies, the autonomy of exploration vehicles will become increasingly critical. The ability to process data onboard eliminates the dependency on continuous communication with ground control, significantly reducing delays and enabling real-time decision-making. For example, a rover equipped with an efficient segmentation model can identify and navigate around obstacles autonomously, ensuring uninterrupted progress even in the absence of direct oversight.

The flexibility of our proposed pipeline extends beyond rock segmentation. By design, it accommodates a variety of image analysis tasks, from identifying geological features to detecting potential scientific interest points. The modular structure allows researchers to integrate their preferred deep learning architectures, optimize them for edge devices, and test their robustness under simulated extraterrestrial conditions. This versatility ensures that the pipeline can evolve alongside advancements in both machine learning and rover technology.

One of the standout contributions of our work is the systematic quantification of both functional and non-functional capabilities of the models. Metrics such as Dice similarity, inference time, and memory footprint provide a comprehensive view of model performance, ensuring that the selected algorithms meet the stringent requirements of onboard deployment. The inclusion of non-functional metrics is especially critical in resource-constrained environments, where trade-offs between accuracy and efficiency must be carefully managed.

Our results highlight the significant potential of post-training quantization as a compression technique. Models that underwent this process maintained segmentation accuracy comparable to their full-precision counterparts while achieving a reduction in size by a factor of 11. This reduction not only facilitates deployment on edge devices but also opens up possibilities for multitasking, where multiple models can operate simultaneously without overwhelming system resources.

Looking ahead, the deployment of these models in real-world missions will require further refinements. Hardware-in-the-loop testing, where algorithms are evaluated on actual rover systems or high-fidelity simulators, will be a crucial step. Additionally, integrating these models with other subsystems, such as navigation and power management, will ensure a cohesive and efficient operational framework.

The insights gained from this study provide a roadmap for future developments in autonomous systems for space exploration. By addressing the dual challenges of robustness and efficiency, our work paves the way for more ambitious missions that push the boundaries of what is possible in remote and hostile environments. As exploration vehicles become increasingly intelligent and capable, the dream of unlocking the mysteries of distant worlds edges closer to reality.

The outcomes of this study also underscore the importance of developing standardized evaluation frameworks for onboard machine learning systems. The lack of uniform benchmarks for comparing functional and non-functional capabilities of such models often leads to fragmented progress. By providing a comprehensive set of evaluation criteria and ensuring reproducibility through publicly accessible implementations, this work contributes to the establishment of best practices in the field. This standardization is vital not only for advancing the current state of the art but also for fostering collaboration across interdisciplinary teams involved in space exploration.

Moreover, the insights gained from simulating various image acquisition conditions point to the need for resilient model designs. Real-world data from extraterrestrial environments are inherently noisy and often deviate from the ideal conditions under which most machine learning models are trained. The simulation pipeline developed in this study highlights the potential of synthetic data to enhance model robustness. By incorporating noise, blur, and varying exposure levels during training, models can be made more adaptable to the unpredictable nature of onboard imaging conditions. This approach could also be extended to other domains where data acquisition is constrained, such as underwater exploration or disaster response scenarios.

Another significant avenue for future research involves integrating multitask learning into the proposed pipeline. While this study focuses on rock segmentation, many tasks in space exploration, such as object detection, terrain classification, and hazard assessment, share underlying features. A unified model capable of performing multiple tasks efficiently could further optimize resource utilization onboard edge devices. Such a development would align with the overarching goals of reducing mission costs and maximizing scientific output.

From a hardware perspective, the continued evolution of edge computing devices will play a pivotal role in expanding the capabilities of onboard systems. Advances in processing power, memory capacity, and energy efficiency will allow more complex models to be deployed without compromising other critical functions. The proposed pipeline is designed to remain adaptable to these advancements, ensuring its relevance in future mission architectures.

Additionally, the lessons learned from this study could inform the design of next-generation space exploration vehicles. For instance, optimizing the placement and calibration of onboard sensors and cameras to complement the strengths of machine learning models could enhance the overall effectiveness of the system. Collaborative efforts between hardware and software teams will be essential to achieving such seamless integration.

Finally, the broader implications of this work extend to the development of autonomous systems in general. The principles of efficiency, robustness, and adaptability that underpin the proposed pipeline are applicable to a wide range of applications beyond space exploration. Industries such as autonomous driving, precision agriculture, and industrial automation could benefit from adopting similar strategies for deploying machine learning models in constrained environments.

In conclusion, this study demonstrates that through thoughtful design and rigorous evaluation, it is possible to bridge the gap between theoretical advancements in machine learning and their practical application in extreme conditions. By addressing the unique challenges of space exploration, this work not only advances the field of autonomous systems but also inspires new possibilities for innovation in related domains. As we continue to explore the cosmos, the tools and methodologies developed here will undoubtedly play a crucial role in shaping the future of intelligent exploration systems.

**Diego Echeverria-Rios, Peter L. Green** (*School of Engineering, The University of Liverpool, UK*)

## **GAUSSIAN PROCESSES FOR PREDICTING PRODUCT QUALITY IN MANUFACTURING.**

**Abstract:** *This paper presents an AI-driven solution designed to predict product quality in continuous manufacturing processes. The solution integrates process parameters and product quality data collected from the production line. The research is motivated by real-world challenges, focusing on an application within a key industry in the UK. It outlines five essential criteria that any AI solution should ideally meet for continuous manufacturing: scalability, modularity, reliable out-of-sample performance, uncertainty quantification, and resilience to unrepresentative data. The paper examines the limitations of current AI methods based on these criteria and introduces a novel solution. This approach utilizes a generalized product-of-experts Gaussian process, with a noise model built using a Dirichlet process. The model's ability to meet the outlined criteria and its effectiveness in the case study within the foundation industry are thoroughly demonstrated.*

**Keywords:** *ai-driven solution, product quality, continuous manufacturing, process parameters, production line, scalability, modularity, out-of-sample performance*

### **Introduction.**

This paper focuses on a foundation-industry application where a product is produced using a continuous manufacturing process. In this context, data on various process parameters is collected at multiple stages of the production line, while product quality, specifically the fault density (the number of faults per unit area), is measured at the end of the line. The objective of this research is to develop an AI-based model, also referred to as a data-driven or machine-learned model, that can predict the product quality based on the process parameters. Subsequently, this model can be utilized to optimize the manufacturing process.

Drawing from previous studies [1-4] and the authors' extensive experience in developing AI solutions for continuous manufacturing, several key criteria must be considered during the initial stages of model development. These criteria are:

1. **Scalability:** As a continuous manufacturing process generates an ever-increasing amount of data, the AI model must be capable of handling large datasets.
2. **Modularity:** Given that data evolves over time—affected by changes in operations or product variations—older data may become less relevant. The model should be able to easily exclude or update outdated data without disrupting its overall functionality.
3. **Stable out-of-data performance:** The model must be able to handle situations where the data falls outside the range it was trained on. This could occur if a sensor malfunctions or if a new product is introduced, and the model should indicate when it is inappropriate to rely on its predictions.
4. **Uncertainty quantification:** The model should provide a measure of confidence in its predictions, an essential feature when using AI for decision-making.
5. **Robustness to unrepresentative data:** It is crucial that the data used to train the model reflects the true nature of the process. Data anomalies caused by external factors such as equipment maintenance or unmeasured raw material changes must be identified and excluded from training.

The paper further explores how existing AI methods measure up against these five criteria and identifies their shortcomings. The proposed solution utilizes a generalized Gaussian process (GP) model, enhanced with a Dirichlet process to account for noise. The method's performance in meeting these criteria, as demonstrated through a case study in the foundation industry, is detailed and analyzed.

focuses on Gaussian Processes (GPs), an influential machine learning technique employed in both classification and regression problems, with the paper concentrating on its application in regression.

In GP regression, a probabilistic method is used to infer an underlying function based on observed data. A key strength of GPs is their ability to quantify uncertainty in predictions. Unlike traditional methods that infer specific parameters of a function, GPs derive samples from a distribution over functions, making them non-parametric. This means they do not rely on a particular family of regressors, allowing them to flexibly model complex relationships in data. This characteristic of GPs has led to their successful application across various fields such as traffic flow, engine modeling, structural dynamics, and robotics.

Standard GPs operate under the assumption that training data is contaminated by noise following a Gaussian distribution. However, as highlighted by several studies, this assumption often does not hold in real-world scenarios. For example, when data contains noise from non-Gaussian distributions, standard GPs may produce inaccurate predictions. Several approaches have been developed to address this, such as using a student-t likelihood to ignore outliers. These modifications allow GPs to handle non-Gaussian noise, though they come with challenges in terms of computational tractability. Techniques like Markov Chain Monte Carlo (MCMC), variational methods, and Laplace approximations have been introduced to estimate model parameters in these cases. Models designed to handle outliers are sometimes referred to as "robust GPs," which incorporate these advanced likelihood functions to mitigate the influence of noisy data.

The standard GP regression typically assumes Gaussian noise, but this assumption was found to be inaccurate for predicting product quality in the case study of interest. In this context, the goal is to predict fault density in products manufactured through a continuous process, and the noise corrupting the data is not Gaussian.

To overcome this, the authors introduce a Gaussian Mixture Model (GMM) for the observations. This model assumes that the noise corrupting each observation comes from one of several Gaussian distributions. The model uses 1-of-K allocation variables to associate each observation with one Gaussian distribution from the mixture. This allows the noise characteristics to vary depending on the source of the data, making it more flexible and better suited to the real-world manufacturing scenario.

In this approach, the allocation variables are treated as latent variables and are integrated into the model using a Bayesian framework. The model includes priors for these allocation variables, which are updated iteratively as new data is observed. The resulting observation model is a mixture of Gaussian distributions centered around the latent function, which is the actual output of the manufacturing process being predicted (fault density).

This modified approach preserves the closed-form solutions associated with standard GP models while allowing for non-Gaussian noise. The mixture of Gaussians helps to capture different noise processes that affect the product quality measurements, such as sensor errors or external disturbances, which would not be properly handled by a standard GP with Gaussian noise.

Through the Dirichlet Process (DP), the authors are able to model the number of mixture components in a non-parametric way. This allows the model to automatically adjust the number of Gaussian components based on the data, without having to pre-define this number. This flexibility is essential for industrial applications where the noise characteristics may change over time.

In essence, this section outlines how a more sophisticated, non-Gaussian noise model can be integrated into a Gaussian Process framework to improve prediction accuracy in manufacturing processes. This method improves upon the standard GP model by introducing a mixture of Gaussians and using a Dirichlet Process to determine the noise model dynamically.

The paper describes the implementation of the proposed robust Gaussian Process (GP) model, which aims to predict product quality in a continuous manufacturing process by incorporating a mixture of Gaussian distributions for noise modeling. The process involves two main steps: clustering and regression.

**Clustering Step:** The first step involves clustering the residuals using a Dirichlet Process (DP) mixture of Gaussian distributions. A DP is a stochastic process that helps define a distribution over an infinite number of potential components, which is essential for clustering data points. The clustering is based on a mixture of Gaussian components, where each observation is associated with

a specific Gaussian based on its noise characteristics. The clustering is done in a Bayesian framework, where priors are set on the allocation variables (which determine the association of each observation to a particular Gaussian) and the variance parameters of the Gaussians themselves.

**Regression Step:** After clustering, the regression step estimates the latent function using the GP approach. Only the observations that are identified as being corrupted by the least noisy Gaussian are used for this step. This ensures that the regression step is not influenced by outliers or noisy data, leading to more accurate predictions. The GP regression process uses the standard formulation to estimate the parameters that best fit the data, providing predictions of the underlying product quality.

For large-scale applications, the proposed model uses a Product-of-Experts (PoE) approach to scale the computations. Each expert model is trained on a subset of the data, and the predictions from each expert are aggregated to form the final predictive model. This approach allows the model to handle larger datasets without running into memory issues and ensures that the computational cost remains manageable.

## Conclusions

In summary, the model combines the clustering of noisy data using a Dirichlet Process with GP regression, making it scalable, robust, and modular. This allows it to effectively predict product quality in continuous manufacturing processes, addressing challenges like noise, data outliers, and scalability with large datasets.

The comparison is based on synthetic data corrupted by noise sampled from a mixture of Gaussian distributions. The synthetic dataset comprises 150 realizations of a function, and the models' predictive accuracy is evaluated by calculating the Root Mean Squared Error (RMSE).

The experiment uses 150 data points where noise follows a Gaussian mixture model with three components, each having different standard deviations and proportionalities. The RMSE results from both Approach 1 and Approach 2 show that the DPGP outperforms both the standard GP and RGP models. Specifically, the DPGP achieves the lowest RMSE, demonstrating its superior ability to handle noise and outliers. The standard GP, in contrast, struggles with noisy data, leading to higher RMSE values.

The comparison between the models also includes visual analysis, where the DPGP model's clustering outcomes and predictive means are shown. The DPGP correctly identifies noise components and does not overfit the data, unlike the standard GP, which overfits by following outlier data points. The results suggest that the DPGP is not only more robust to outliers but also provides more accurate predictions for unseen data.

In conclusion, the experiment illustrates the DPGP's strong performance, particularly in environments with noisy data, while also showcasing its scalability and ability to integrate new data seamlessly. This comparison highlights the advantages of the DPGP approach in predictive tasks involving complex data, such as in continuous manufacturing processes.

## REFERENCES

1. Liu, H., Cai, J., Wang, Y., Ong, Y.S., 2018a. Generalized robust Bayesian committee machine for large-scale Gaussian process regression. In: 35th International Conference on Machine Learning, ICML 2018, Vol. 7. pp. 4898-4910.
2. Feng, Z., Li, D., Qin, G., Liu, S., 2009. Effect of the flow pattern in a float glass furnace on glass quality: Calculations and experimental evaluation of on-site samples. *J. Am. Ceram. Soc.* 92 (12), 3098-3100.
3. Feng, Z., Wang, J., Ma, Y., Tu, Y., 2021. Robust parameter design based on Gaussian process with model uncertainty. *Int. J. Prod. Res.* 59 (9), 2772-2788.
4. Jin, D., Xu, S., Tong, L., Wu, L., Liu, S., 2020. A deep learning model for striae identification in end images of float glass. *Trait. Signal* 37 (1), 85-93.

**Dominik Pieczyński, Bartosz Ptak, Marek Kraft, Mateusz Piechocki, Przemysław Aszkowski**  
(Poznań University of Technology, Institute of Robotics and Machine Intelligence, Piotrowo 3A,  
Poznań, Poland)

## **FAST, LIGHTWEIGHT DEEP LEARNING PIPELINE FOR UAV LANDING SUPPORT**

**Abstract:** *Despite significant advancements in aerial robotics, achieving precise and autonomous landing in various environments remains a complex challenge. Several factors, such as terrain characteristics, weather conditions, and obstacles, influence this process. This paper introduces a deep learning-based image processing pipeline that accelerates the detection of landing pads and the estimation of the UAV's relative pose. Additionally, the system enhances safety and reliability by incorporating human presence detection and error estimation for both the identification of landing targets and pose computation. Human detection and landing pad localization are achieved through a segmentation-based approach, which estimates the probability of presence. The landing pad's keypoints are then identified using a regression algorithm that not only determines the coordinates but also provides an uncertainty measure for each landing pad feature. To address these tasks, a set of lightweight neural network models was carefully selected and tested. The paper presents performance metrics for each component of the system as well as for the entire processing pipeline. These evaluations, conducted on embedded UAV hardware, demonstrate that the method can deliver precise, low-latency feedback, ensuring safe landing operations.*

**Keywords:** *neural networks, performance metrics, low-latency, embedded hardware, precision, reliability, feedback, localization*

### **Introduction.**

The process of autonomous landing for unmanned aerial vehicles (UAVs) remains a significant challenge, even amidst the rapid advancements in aerial robotics. Several factors contribute to the complexity of this task, including the terrain's characteristics, varying weather conditions, and the potential presence of obstacles in the landing zone. This paper outlines a deep learning-based image processing pipeline designed to enhance the accuracy of detecting landing pads and estimating the UAV's relative pose with respect to these pads. Furthermore, the system incorporates mechanisms for increasing safety and reliability, such as detecting the presence of humans and estimating potential errors in landing target detection and pose computation.

The system utilizes a segmentation approach to estimate the likelihood of human presence and the position of the landing pad. This information is followed by a regression algorithm that identifies key points on the landing pad, determining not only their coordinates but also the associated uncertainty of the detected key points. To achieve these tasks, a set of lightweight neural network models was chosen and thoroughly tested. Performance and accuracy measurements of the system's components are provided, showing that the system can deliver precise, real-time feedback, ensuring safe landing operations.

### **Related Work**

The challenges in autonomous UAV landing can generally be categorized into two approaches: landing on natural surfaces and landing on prepared landing sites, each presenting its own set of difficulties. Landing on natural, unmodified terrains typically involves finding a flat and stable area of appropriate size, a task that often relies on aggregated measurements converted into a cost function. However, this approach tends to be more complex than landing on predefined, feature-rich artificial surfaces. As a result, natural surface landings are not typically considered for high-precision operations, such as landing on moving platforms. Therefore, the system presented in this work focuses on landing on an artificial, helipad-like surface, where the challenges are more manageable due to predefined markings.

Vision-based methods for landing on platforms with defined shapes and textures commonly utilize keypoints located on the surface to compute the relative pose between the UAV and the landing pad. The use of a sparse set of keypoints for pose estimation has been widely recognized as a reliable technique in this context. For instance, in the early works by [1], elementary image processing operations like image binarization, segmentation, and feature detection were used to identify



keypoints on a template marker. Though simple by today's standards, this approach laid the foundation for more advanced techniques.

As the field evolved, more sophisticated vision-based systems were developed. These systems often relied on hand-designed patterns or general-purpose markers such as QR codes, ArUco codes, or AprilTags for precise pose estimation. Some methods, like those proposed by [2] or [3], used shape analysis and the fitting of ellipses to landing pad templates, in combination with IMU data, to estimate the UAV's relative pose. Other systems, such as [4], employed patterns involving concentric circles. These hand-designed markers could also incorporate color information, like the marker described by [5], which used color segmentation and Hu's moments for reliable operation in cluttered environments.

Additionally, the application of active markers, which emit infrared radiation detected by thermal cameras or infrared sensors, provides another layer of precision. While these methods offer advantages in terms of accuracy, they require additional power supplies and specialized sensors, which may not always be feasible in UAV applications.

The rapid development of deep learning has further advanced vision-based landing systems. Modern approaches, such as the ones used by [6], incorporate convolutional neural networks (CNNs) for detecting landing pads. These systems enhance detection accuracy and can handle situations where the landing pad is partially obstructed, although they are often computationally intensive and do not compute the relative pose directly.

Lastly, recent work has focused on real-time systems that use deep neural networks for person detection and safe landing zone estimation. Overall, while many of these methods offer solutions for specific aspects of the landing process, they often lack full integration or fail to meet the high precision and safety requirements needed for autonomous UAV landings. This paper proposes a more holistic approach that not only detects the landing pad but also ensures the safety of the landing environment, incorporating real-time feedback to adjust landing maneuvers dynamically.

The complete pipeline consists of multiple components: a camera, image processing operators, and a control algorithm for the UAV's platform. The camera captures images, which are then processed by a set of lightweight neural networks designed to handle complex operations. These networks detect and analyze objects like the landing pad and human presence, providing real-time feedback.

The segmentation model processes the input image and generates two probability masks, one for detecting human presence and another for the landing pad. This segmentation is followed by contour extraction to identify regions of interest. When human presence is detected within the threshold distance from the landing pad, the landing process is halted for safety reasons. If the landing pad is detected, the algorithm identifies the largest region of interest (ROI) and applies a keypoint detection model. This model estimates the locations of keypoints on the landing pad, as well as the coordinates' variance and confidence level.

The keypoint regression step then uses the keypoint information to calculate the relative pose of the UAV to the landing pad. This is done using a 3D model of the landing pad and a perspective-N-point solver. The resulting pose is translated into the global coordinate system, allowing the UAV to adjust its position accurately.

This methodology is designed to be implemented on embedded edge AI devices, optimizing the system's computational performance while ensuring real-time operation for precise landing, even under challenging conditions. The choice of neural networks and image processing techniques is based on their ability to operate efficiently on resource-constrained hardware. The use of lightweight models allows for high-throughput, real-time processing suitable for UAV applications.

We describe the hardware setup used for benchmarking the algorithm and introduce the external simulation environment via Robot Operating System (ROS). This external simulation enables early-stage algorithm testing in a virtual environment while using target hardware, following the hardware-in-the-loop approach. This allows both synthetic and real-world data to be fed to the same hardware controller.

#### **4.1 Flying Platform**

The research was conducted on a Holybro X500 quadcopter, a commercially available flying platform selected for its open design and modification flexibility. This platform is powered by the Pixhawk 4 autopilot flight controller running PX4 firmware version 1.11.3. The system is open-source and well-documented, allowing easy adaptation for various applications. The platform's carbon frame, which weighs only 1000 g without batteries, supports four 2216 KV880 motors and  $10 \times 4.5$  propellers, powered by a 5000 mAh Li-Po 4s battery. With a maximum take-off weight of 2000 g, the platform can carry additional payloads, such as cameras and onboard computers, while maintaining a flight endurance of over 10 minutes.

#### 4.2 Hardware Description

To ensure the onboard processing capabilities for the UAV, we chose small, lightweight devices with low power consumption. The goal was to perform neural network inference on edge AI devices with minimal energy requirements. We evaluated the following embedded AI devices:

- **Nvidia Jetson Nano:** This device is the lowest-performance model in the Jetson family, delivering 472 GFLOPs while consuming just 10 watts. It features a quad-core Cortex A57 ARM processor and a Maxwell GPU with 128 CUDA cores, weighing only 140 g.
- **Nvidia Jetson Xavier NX:** A more powerful device designed for embedded applications, it includes a six-core CPU, 384 CUDA cores, and 64 tensor cores. It also contains two Nvidia Deep Learning Accelerator (NVDLA) engines for deep learning tasks, with a weight of 172 g.

For the vision sensor, we used the **Luxonis Oak-D-Lite**, a USB camera with a color sensor that provides full-HD frames and a 81-degree field of view. This camera was chosen for its low cost and versatility, and although it also captures depth images via a stereo pair, the depth information was not used in these experiments.

#### 4.3 Drone Simulation and Off-Board Control

Recent advancements in flight controllers, such as PX4, allow for control via external signals. This is made possible by integrating a companion computer on board, which provides top-level control commands with a minimum frequency of 2 Hz. To override the autopilot control, the flight controller must be in off-board mode. Communication between the systems is established through the MAVROS node, which translates ROS topics into MAVLink-compatible messages.

Off-board control of UAVs requires rigorous testing to avoid dangerous situations that could cause the vehicle to crash. To ensure safety, we used both software-in-the-loop (SITL) simulations with PX4 and real-world hardware for testing. The simulation runs on a personal computer, while communication with the embedded device is handled over a local network, enabling the algorithms to be tested directly on the target hardware without additional delays.

Finally, the entire inference pipeline was evaluated, taking into account both speed and accuracy. For segmentation, the best performance was achieved using the DeepLabV3Plus + ResNet10t combination, while for keypoint regression, the MobileNetV3\_100 and LCNet\_050 models were selected due to their higher processing speeds. Aggregating the results, the system achieved  $12.86 \pm 0.11$  FPS on the Jetson Nano and  $83.38 \pm 6.96$  FPS on the Jetson Xavier NX. The best speed-performance combination was found with DeepLabV3Plus + LCNet\_050 for segmentation and MobileNetV3\_100 for keypoint regression, which resulted in  $19.33 \pm 0.51$  FPS for the Jetson Nano and  $93.64 \pm 1.22$  FPS for the Xavier NX.

It is important to note that the vision pipeline must meet specific computational requirements to work effectively. The use of deep learning necessitates that the UAV platform be equipped with an embedded GPU to perform inference at the required speed, limiting its use to UAV platforms that have the necessary space and power capacity to carry additional devices and their power supply.

Further experiments showed that the pipeline is capable of detecting the landing target and keypoints even under challenging conditions, such as motion blur, scaling, perspective changes, and poor lighting. The model's robustness was evaluated with data augmentation techniques like perspective transformations, scaling, rotation, and the addition of noise. The results confirm that the algorithm can handle these real-world conditions. Additionally, it can fill in missing keypoint data caused by occlusions or damage to the landing pad, while also informing the system about the

uncertainty in keypoint positions. This feature is crucial for making informed decisions about whether or not to proceed with landing, contributing to a more cautious approach when necessary.

The authors present a deep neural network-based vision pipeline designed to offer precise landing support for UAVs. This system enables real-time processing, delivering landing target information, including positional and rotational data relative to the UAV's camera, ensuring precise landings. The method is optimized to run efficiently on embedded GPU devices commonly used in UAV platforms, processing up to 20 frames per second.

One of the key differences between this approach and others is the use of deep learning algorithms to not only detect landing pads but also handle potential human presence in the landing zone. Additionally, the system assesses the confidence in the detected keypoints' positions, which further enhances landing safety. The method is resilient to environmental factors such as lighting variations, scaling, and rotation, demonstrating robustness even when dealing with partial data or occlusions, thanks to its advanced explainability techniques.

The relative pose estimation's accuracy was evaluated using a high-precision external system, confirming that the deep learning model can estimate landing pad keypoints with sufficient accuracy for reliable landing. However, the system's adaptability to various landing pads depends on their visual characteristics, such as high contrast with the environment and the presence of at least four distinct, non-collinear keypoints.

Future research will focus on refining this system for real-world UAVs, especially for landing on moving platforms, which introduces additional challenges. The authors also intend to explore optimization techniques to improve processing speed, such as hardware upgrades and neural network optimization methods like quantization and pruning. This future work aims to ensure that the system can handle even more complex landing scenarios.

## REFERENCES

5. Sharp, C.S., Shakernia, O., Sastry, S.S., 2001. A vision system for landing an unmanned aerial vehicle. In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, Vol. 2. Ieee, pp. 1720-1727.
6. Zhang, G., Zhou, F., 2005. Position and orientation estimation method for landing of unmanned aerial vehicle with two circle based computer vision. *Acta Aeronaut. Astronaut. Sinica* 26 (3), 344-348.
7. Yang, S., Scherer, S.A., Schauwecker, K., Zell, A., 2013a. Onboard monocular vision for landing of an MAV on a landing site specified by a single reference image. In: *2013 International Conference on Unmanned Aircraft Systems. (ICUAS), IEEE*, pp. 318-325.
8. Lange, S., Sunderhauf, N., Protzel, P., 2009. A vision based onboard approach for landing and position control of an autonomous multirotor UAV in GPS-denied environments. In: *2009 International Conference on Advanced Robotics. IEEE*, pp. 1-6.
9. Garcia-Pulido, J., Pajares, G., Dormido, S., 2022. UAV landing platform recognition using cognitive computation combining geometric analysis and computer vision techniques. *Cogn. Comput.* 1-21.
10. Lee, S., Jo, D., Kwon, Y., 2022a. Camera-based automatic landing of drones using artificial intelligence image recognition. *Int. J. Mech. Eng. Robotics Res.* 11 (5), 357-364.

**Federico Lombardo, Federico Pittino, Daniele Goldoni, Luca Selmi** (*University of Modena and Reggio Emilia, Modena, Italy*).

## MACHINE LEARNING FOR NANOPARTICLE SIZING WITH BIOSENSOR ARRAYS

**Abstract:** *Many technology applications still exist where Artificial Intelligence techniques, carefully tailored to meet specific application requirements, can deliver significant improvements in hardware performance. One such area is biosensing using innovative complementary-metal-oxide-semiconductor (CMOS) nanocapacitor arrays. These sensors have the potential to function as advanced imaging platforms, but despite the progress in the field, there remains a significant gap in the knowledge required to accurately and reliably interpret their responses to analytes. This study harnesses the power of Machine Learning (ML) methods in computer vision to develop precise and robust models across various operational conditions. By recognizing the parallels between multifrequency capacitance maps and multispectral images, we identify the most suitable ML algorithms to predict the size of analytes detected by the nanoelectrode array biosensor. As a key example, we focus on measuring the radius of dielectric spherical nanoparticles dispersed in deionized water and phosphate-buffered saline. We compare the performance of large, well-established image-processing neural networks to simpler, purpose-built models. Large training datasets are generated by combining precise finite element simulations of sensor responses with experimental data. Remarkably, we achieve an excellent level of accuracy, comparable to traditional sizing methods, when determining the size of nanoparticles that are similar in scale to the pixel pitch of the array. The median size error remains below 15% across all scenarios when a small fraction of measured data samples is added to the simulation-based training dataset.*

**Keywords:** *artificial intelligence, machine learning, biosensing, nanocapacitor arrays, capacitance maps, multispectral images, nanoparticles, simulation-based training*

### Introduction.

Machine Learning (ML) and Deep Learning (DL) techniques have revolutionized many fields, including computer vision, speech recognition, natural language processing, and generative models. However, their application in certain sectors is still in its early stages, despite the substantial potential they hold. One such area is biosensing, particularly when it comes to accurate and reliable measurements of biological objects suspended in liquid solutions, such as biomolecules and cells. Although there have been some attempts to integrate ML into biosensing technologies, including limited applications for biosensors, significant challenges remain. A primary obstacle is the lack of extensive, high-quality datasets that can effectively cover the entire operational space of these sensors. This scarcity of data hinders the ability to achieve the high accuracy and robustness that ML-based systems typically promise.

One promising direction for improving the accuracy of conventional biosensors is to incorporate ML-based data analysis. This approach has the potential to enhance both the precision of measurements and the reliability of data interpretation. However, the presence of noise, variability, and uncertainty in measurements complicates the process, often making accurate and cost-effective data analysis difficult. Furthermore, accurate physics-based simulations, which could provide cleaner and larger datasets for training, are still not widely available. Additionally, most biosensing platforms operate with limited parallelism, further restricting the collection of large datasets across multiple analytes. In recent years, complementary metal–oxide–semiconductor (CMOS) micro/nano-electrode array systems have been developed to address some of these challenges. These systems offer various advantages, such as miniaturization, high parallelism, and the ability to mitigate the spatial sensitivity limits imposed by Debye screening. These biosensors can detect nano-scale objects (objects with dimensions  $\leq 1 \mu\text{m}$ ) without requiring labels. As a result, CMOS-based nanoelectrode array biosensors have gained attention for applications ranging from measuring microparticles and nanoparticles to detecting pollutants in environmental monitoring.

While these systems show promise, they still require further optimization, particularly when it comes to accurately estimating the size of nanoparticles at resolutions that match the pixel pitch of the sensor arrays. This work explores how advanced ML methods, originally designed for computer vision tasks, can be adapted for estimating nanoparticle sizes based on multi-frequency capacitance data collected by nanoelectrode array biosensors.

By leveraging the structural similarities between capacitance maps and multispectral images, this study identifies suitable ML algorithms to predict the size of nanoparticles, specifically focusing on spherical dielectric particles. The results suggest that machine learning methods, when appropriately trained on simulations and real-world measurements, can achieve high accuracy in nanoparticle sizing, even when the particle sizes approach the resolution limits of the sensor arrays. This work paves the way for more accurate, cost-effective biosensing technologies capable of monitoring and quantifying nanoparticles in complex environmental and biological systems.

### **1 Measurements: Experimental Setup, Procedure, and Data Preparation**

The work utilizes a High Frequency Impedance Spectroscopy (HFIS) nano-electrode array chip, which was designed and provided by NXP Semiconductors. This CMOS chip incorporates a 256×256 array of individually addressable nano-electrodes (NEs), each with a 90 nm radius and a 600 nm × 720 nm pitch, fabricated with 90 nm CMOS technology. The chip uses Charged-Based Capacitance Measurement (CBCM) and ad hoc calibration procedures to measure capacitance within a frequency range of 1-70 MHz (with potential extension to 500 MHz).

Capacitance images are generated by sequentially activating one row of electrodes at a time while grounding the others, effectively creating a large counter electrode. The capacitance change due to analyte arrival is measured by the difference in capacitance before and after the analyte's interaction with the electrodes. Polystyrene (PS) nanoparticles of specific sizes (275 nm and 500 nm nominal radii) suspended in MilliQ water and PBS were used for the measurements. The nanoparticles were prepared using a combination of dynamic light scattering (DLS) and filtering to ensure purity. The system was initialized, allowed to reach thermal equilibrium, and measurements were conducted at 22°C.

Data preprocessing involved the identification of capacitance maps (7×7 electrodes) that contained only a single nanoparticle. Noise reduction strategies were implemented to eliminate outlier data and match experimental results with simulations. Specific steps were taken to calculate the mean and standard deviation of capacitance from the outermost electrodes, and a noise model was introduced to simulate realistic measurement conditions.

### **2 Numerical Simulations: Calculations and Data Preparation**

Simulations of the nanoelectrode array response were conducted using the ENBIOS software, which solves the Poisson–Boltzmann and Poisson–Drift–Diffusion equations to calculate capacitance changes due to nanoparticles. Simulations were performed under two conditions: one with the particle present and one without, across a matrix of electrodes surrounding the particle. For accurate simulations, only the innermost 7×7 electrodes were considered.

Simulations were performed for the MilliQ environment using a constant frequency of 50 MHz, and for PBS using a frequency range of 2–70 MHz. The simulated data included various particle sizes (ranging from 200 nm to 1500 nm) and positions, with an augmentation technique involving flipping images to increase model robustness.

### **3 Radius Estimation Model Development**

To estimate the particle radii, machine learning (ML) algorithms were applied to the multifrequency capacitance maps, taking advantage of the similarity to regular images used in computer vision. Several deep learning models, including MobileNet-V3Small, ResNet18, and simpler convolutional neural networks (CNNs), were developed and tested. Special emphasis was placed on creating smaller models to prevent overfitting, due to limited training samples and the need for low-memory, low-power applications in edge devices.

Two variants of the MobileNet architecture, Tiny MobileNet and SuperLite MobileNet, were created with reduced parameters for optimal performance in this application. Additionally, CNN architectures, such as the Two Blocks Network and Simplest Network, were designed for their simplicity and effectiveness in nanoparticle sizing.

The models were trained using the Gaussian Negative Log-Likelihood loss function, which allows the model to estimate both the mean and variance of the particle radius, aiding in uncertainty quantification. The training was performed using the Adamax optimizer.

### **Results in MilliQ Environment:**

### **Training and Test on Simulated Samples Only**

Initially, the ML models were trained using only simulated data. The training data set included all nanoparticle radii except the 750 nm samples. White Gaussian noise, estimated from the measurement data, was added to the simulation data to stabilize the results and reduce model overfitting. The hyperparameters of the models were optimized on a validation set, and their performance was tested on a separate test set of simulated particles with a radius of 500 nm.

The results, as depicted in the figures, reveal that models with fewer parameters, such as the Simplest Net and Two Blocks Net, exhibited the best performance, achieving median errors below 10%. In contrast, larger models like MobileNet showed severe overfitting, which significantly degraded their performance, thus excluding them from further analysis.

### **Test on Measurement Samples (Different Trainings)**

Following the training on simulation data, the models were also tested using real measurement samples of nanoparticles with radii of 275 nm and 500 nm. The training data included only a small subset of measurements, and the models were evaluated based on how well they generalized to these new samples. Notably, Tiny MoNet, which had performed poorly in previous stages, was excluded from this part of the analysis.

Results demonstrated that, except for Tiny MoNet, all models successfully distinguished between the two nanoparticle radii. The Simplest Net exhibited the lowest median error and the least spread in its predictions. However, when a small number of measurement samples (10 per radius) were included in the training data, performance did not significantly improve for most models, except for the SuperLite MoNet, which showed enhanced accuracy. The lack of improvement was attributed to the sensitivity of the measurements to uncontrolled variables, such as particle displacement and alignment relative to the electrode, particularly for nanoparticles smaller than the electrode pitch.

### **Results in PBS Environment:**

#### **Training and Test on Simulated Samples Only**

The same procedure as in MilliQ was applied to the PBS environment. The results confirmed previous findings: models with fewer parameters, like Simplest Net and Two Blocks Net, demonstrated superior accuracy, with median errors of only a few percent. SuperLite MoNet, on the other hand, performed poorly, exhibiting a higher spread of errors, especially for the larger nanoparticles (500 nm).

#### **Test on Measurement Samples (Different Trainings)**

The models were then tested on measurement data in PBS. Several adjustments were made in the training process to account for the effects of salinity and multi-frequency data. The accuracy of noise estimation was improved by only considering measurements from the 275 nm radius nanoparticles. A data augmentation technique was applied, adding random Gaussian noise to the ground truth values of the radius during training to improve model robustness.

The results from testing on the PBS measurement data showed a significant improvement when a small subset of measurement samples was added to the training set. In particular, SuperLite MoNet, which had performed poorly on the simulation-only training set, became the most accurate model after incorporating measurement data.

### **Summary of Results:**

In summary, the models trained on simulations with a small number of measurement samples showed the best performance, providing reliable nanoparticle sizing even when the particle size was close to the pixel pitch of the array. The Simplest Net and SuperLite MoNet models achieved the highest accuracy, with median errors of less than 15% in all scenarios, including tests in both MilliQ and PBS environments.

In contrast, relying solely on measurement data proved less effective, particularly for the larger nanoparticles in PBS. The addition of simulated data greatly enhanced model performance by increasing the diversity of the training set and providing more stable predictions. These results underscore the potential of combining simulated data with real measurements for training robust machine learning models, which is crucial for nanoparticle metrology using advanced biosensor technologies. This study provides a significant contribution to enhancing the functionality of

nanoelectrode array biosensors with the integration of optimized machine learning (ML) algorithms. These algorithms are applied to the capacitance measurements of nanoparticles in liquid solutions, specifically focusing on estimating their physical parameters, such as radius. The work exploits the structural similarities between multi-frequency capacitance maps obtained from the sensors and regular multi-spectral images. This approach enables the application of advanced ML techniques traditionally used in computer vision. The case study centers around the estimation of nanoparticle radii, a particularly challenging task due to the sub-micron size of the particles and the smaller dimensions compared to the pixel pitch of the imaging system. The study utilizes simulated data as an augmentation technique to improve model performance, enhancing the metrology of nanoparticles. The primary findings of this research include the successful application of simplified versions of state-of-the-art deep learning models originally developed for RGB image recognition. These models have been tailored to operate with the specific characteristics of capacitance maps. Additionally, the study demonstrates how machine learning can overcome some of the current limitations of nanoparticle metrology, such as detecting and sizing particles in aqueous environments without the need for complex and costly optical techniques. This is especially relevant in the context of monitoring nanoplastics in water, where current methods are inadequate.

A crucial aspect of the methodology is the incorporation of simulated data samples to train the ML models, which proves to be an effective strategy in enhancing the accuracy and robustness of nanoparticle measurements. The models trained with this data achieve performance comparable to traditional dynamic light scattering (DLS) methods, but with the added benefit of being able to operate in a label-free, high-throughput manner. Overall, the research shows promising results, indicating that nanoelectrode array biosensors, when coupled with optimized machine learning models, can provide accurate, real-time measurements of nanoparticle characteristics, facilitating future advancements in fields such as environmental monitoring and medical diagnostics.

**Federico Rossi, Cinzia Bernardeschi, Marco Cococcioni** (*University of Pisa, Pisa, Italy*)

## **NEURAL NETWORK VERIFICATION IN CLOSED-LOOP SYSTEMS USING INTERVAL ARITHMETIC).**

***Abstract:** Machine learning methods have been effectively applied to develop high-performance control components for cyber-physical systems, where control dynamics arise from the integration of multiple subsystems. However, these methods may not always provide the level of trustworthiness needed for safe operation in critical environments. This paper presents an innovative approach that combines interval arithmetic with theorem-proving techniques to verify safety properties in closed-loop systems incorporating neural network components. We apply this approach to a model-predictive controller for autonomous driving, comparing the verification results of the neural network with those obtained using other existing tools. Our findings reveal that open-loop verification of neural networks using interval arithmetic can outperform current methods, achieving property proofs with reduced time overhead. Additionally, we demonstrate how the combination of these two techniques allows for constructing a formal model of the network in higher-order logic, applied to the controlled closed-loop system.*

***Keywords:** machine learning, control components, cyber-physical systems, subsystems, trustworthiness, safety-critical, interval arithmetic, theorem-proving*

### **Introduction.**

In recent years, machine learning has played a crucial role in developing high-performance control systems for cyber-physical systems (CPS). These systems, which combine computational algorithms and physical components, often operate in a closed-loop setup where control dynamics emerge from multiple interacting subsystems. While machine learning approaches, especially neural networks, have significantly improved the capabilities of CPS in fields like autonomous vehicles, smart manufacturing, and medical applications, they come with challenges. One of the key concerns is their trustworthiness in safety-critical environments, where any malfunction could have severe consequences.

To address these concerns, we propose combining two verification techniques: interval arithmetic and formal verification methods. This approach helps assess the safety properties of closed-loop systems that use neural network controllers. Specifically, interval arithmetic is used to bound the possible outputs of the neural network, ensuring that the system behaves within safe limits under all possible inputs. Formal verification, on the other hand, uses logical proofs to guarantee that the system meets safety requirements, offering a higher level of assurance.

We demonstrate the effectiveness of this combined approach by applying it to a model-predictive controller for autonomous driving. Our results show that this method can verify neural network properties more efficiently than existing tools, with less computational overhead. Moreover, by integrating both techniques, we can create a formal model of the neural network within a higher-order logic framework, making it easier to prove the safety of closed-loop systems.

In the future, this method could be applied to a wide range of critical systems, ensuring that machine learning-based controllers are not only powerful but also safe and reliable for real-world applications.

### **2. Current State of the Art**

Recently, significant efforts have been focused on formal or analytical verification of neural networks using various theoretical approaches. Most of these methods are concentrated on verifying networks with ReLU activation functions for multi-layered networks, where the activation function has a linear nature at each stage.

One of the most commonly used methods involves SMT-based tools for verifying neural networks. These tools help check whether the output values of the network meet the specified requirements, ensuring the network's correctness for certain input data.

Another approach is the use of hybrid automata for formalizing networks with more complex activation functions, such as sigmoid or hyperbolic tangent. This approach allows the verification of



not only linear but also nonlinear activation functions, opening new opportunities for modeling more complex neural networks.

Simulation-based methods, particularly Monte Carlo, are also widely used for the empirical analysis of systems with neural networks. These methods simulate the probabilistic behavior of the network for different input values, providing a statistical assessment of the network's stability and accuracy in real-world conditions.

Each of these approaches has its advantages but also limitations, especially in terms of scalability for large networks and complex dynamic systems. In our work, we focus on integrating several approaches to achieve more efficient and reliable verification results for closed-loop systems that embed neural networks as control components.

The third section of the document provides background information on the verification tools and methods used in the paper. It discusses the Prototype Verification System (PVS), which is a mechanized environment for formal specification and verification. PVS enables users to specify systems through theories, which are sets of formulas, variable declarations, and function declarations. The PVS language supports a variety of base types like naturals, integers, reals, and booleans, all of which are defined in a fundamental library prelude. More complex data types, such as matrices, are provided by the Nasalib extensions.

The PVS proof system is based on sequent calculus, which allows users to prove theorems by manipulating sequents. In sequent calculus, a sequent is a logical expression that separates the antecedents (premises) from the consequents (conclusions). The system can perform inference rules to break down complex proofs into simpler ones, and a proof is considered successful when all branches of the proof tree have been validated.

The document then outlines the neural network structure used in the verification approach. It describes the input-output relations for neural networks and introduces a specific type of network architecture, the fully connected feed-forward neural network. The network is defined by a series of layers, each with its own weights and biases. The neural network is mathematically represented in terms of matrix operations, where the activation function plays a key role in transforming the network's output.

Interval arithmetic is used to evaluate the possible output range of a neural network by calculating bounds for its outputs given the range of possible inputs. In the approach outlined in this section, each input feature to the neural network is represented as an interval with upper and lower bounds, and the goal is to compute the output intervals for each neuron in the network. The key idea is to propagate these input intervals through the layers of the network, applying interval arithmetic operations at each layer. These operations evaluate the bounds for intermediate activations, which are then used as input intervals for the subsequent layers. This process continues until output intervals are obtained for the final layer of the network.

The use of interval arithmetic ensures that the computed output bounds are guaranteed, meaning that the neural network will always produce results within the computed intervals for any inputs that fall within the specified input bounds. This provides a systematic and reliable way of verifying neural network behavior, particularly in systems where safety and dependability are paramount.

The section also emphasizes the importance of handling non-linear activation functions, such as the hyperbolic tangent (tanh) and sigmoid, which are commonly used in neural networks. The approach described applies to feed-forward neural networks, especially in regression tasks, and focuses on providing verified safety properties for closed-loop systems controlled by neural networks.

The neural network is first modeled as a sequence of transformations, starting with the input layer and passing through each subsequent fully connected layer. Each layer is represented by a mathematical function in PVS, which operates on matrices to describe the transformation of the network's input through weights and biases, followed by an activation function. This formalization ensures that the neural network's operations are captured precisely, using the formal language of higher-order logic.

Input constraints are also imposed in the formal model, with each input variable bounded within specific ranges. These bounds are essential for ensuring the network's behavior adheres to safety properties. The PVS model also specifies properties of the network's outputs, such as ensuring that they remain within acceptable limits under certain conditions, and these properties are proved within the formal framework.

Finally, the paper discusses how this formal model, once constructed, can be used to prove safety properties of systems where the neural network is deployed in a closed-loop system, such as in autonomous driving or industrial control applications. The neural network's behavior is rigorously analyzed, ensuring that it operates within defined safety bounds in these critical applications. This formal verification process, leveraging PVS and interval arithmetic, provides a solid foundation for the safe deployment of neural networks in cyber-physical systems.

### **1 Automatic Theory Generation**

To address the challenge of scaling up the verification process for large networks, the paper introduces a Python-based tool designed to automatically generate a PVS theory starting from a pre-trained PyTorch neural network model. This process efficiently converts the neural network's model into a formal theory, allowing for automatic verification. The complexity of generating the theory depends on the number of parameters in the network, with the time for creating a theory being relatively short (e.g., two seconds for a simple network with a 3GHz desktop processor). However, certain assumptions are necessary to ensure the theory generation is tractable:

- The network must be a feed-forward model (e.g., a multi-layer perceptron).
- The number of neurons should be limited to 60-70 to maintain scalability with the prover.
- Linear activation functions are recommended for tighter bounds and to reduce over-approximations.

### **2 Interval Arithmetic Computation**

The paper integrates interval arithmetic into PyTorch by extending the base Tensor class to include an IntervalTensor class, where each tensor element represents a range of values. This allows for interval-based computations on neural network models. The pyinterval library, which implements algebraically closed interval operations, is used for this purpose. This extension makes it possible to propagate input intervals through the network, applying interval arithmetic to each layer's output, thus generating bounds for the network's output.

This integration allows the IntervalTensor to function seamlessly with pre-trained models in PyTorch without requiring recompilation of the entire framework. By overriding the layer operation implementations within the IntervalTensor class, it provides a common interface that adheres to PyTorch's original structure, enabling efficient interval arithmetic without disrupting the PyTorch model flow.

### **3 A Small Example of the Two Techniques**

To illustrate the practical application of these techniques, the paper provides an example of a simple neural network with a linear layer followed by a ReLU activation function. The corresponding PyTorch code for this model is provided. Using the automatic theory generation tool, the paper demonstrates how a formal PVS theory can be generated for this simple network. The constraints on the input variables and the corresponding theorem to verify the output are specified, showing how these methods work in practice.

### **4 Limits of the PVS Theory Generation and Proof**

The section also highlights some limitations of the approach. As the size of the network increases, both the complexity of generating the formal theory and the associated computational costs increase. The scalability of the theory generation process is constrained by the size of the network, with the automatic tool being effective for moderately sized networks but encountering challenges with very large ones.

In conclusion, Section 6 focuses on the integration of formal verification tools within the PyTorch framework, showing how interval arithmetic can be applied to neural networks for proving their safety properties in a closed-loop system. This approach is designed to be scalable, although it has practical limits when applied to very large networks.

In this use case, the application of neural networks for controlling an autonomous vehicle's adaptive cruise control system is explored. The scenario considered involves the vehicle (referred to as the "ego car") adjusting its speed and acceleration based on the position and speed of a lead car, which is detected through sensors.

The MPC uses three nominal inputs: the relative velocity of the lead car, the relative distance, and the ego car's longitudinal velocity. The output of the controller is the acceleration of the ego car, which is the only regulated variable.

The paper proposes replacing the MPC with a surrogate model—a neural network—capable of handling the same inputs to generate the desired acceleration output. The neural network is trained on data consisting of input-output tuples: the lead car's relative velocity, relative distance, and the ego car's longitudinal velocity, with the corresponding acceleration of the ego car. The network is trained using the ADAM optimizer for 1000 epochs with a mean squared error (MSE) loss function, achieving an MSE of 0.034.

The substitution of the MPC with a neural network in this context is motivated by its potential for more efficient computation, particularly in embedded systems. The MPC traditionally solves a constrained quadratic programming (QP) optimization problem in real time. However, using a neural network to replace the MPC offers computational advantages, especially when real-time problem-solving is required, as neural networks can provide a faster solution.

This safety property is crucial to verify the reliability of the autonomous system, particularly in the context of a neural network replacing the original MPC.

To evaluate the neural network's performance, the boundaries of the network's output are analyzed. The neural network's outputs are evaluated by adjusting the input parameters, particularly the relative distance between the two cars. This test simulates the scenario where the ego car is too close to the lead car and needs to decelerate to maintain the safety distance.

Comparing the verification times of different neural network verification tools, the results show that the proposed method significantly reduces the time overhead, making it more efficient than existing approaches. This is particularly valuable in the context of safety-critical systems where verification time is a crucial factor.

Through this process, the approach not only ensures that the neural network behaves safely within the defined bounds but also demonstrates how the combination of formal verification techniques, such as interval arithmetic and theorem-proving, can be effectively used to guarantee the reliability of neural networks in dynamic and closed-loop systems, such as autonomous driving.

## **Conclusion**

We have developed an innovative verification method that combines the advantages of interval arithmetic with formal verification through theorem proving. This hybrid approach offers a strong framework for validating the safety and dependability of neural networks in closed-loop systems.

Our method is fully compatible with the PyTorch deep learning framework, allowing for easy integration without requiring significant modifications to existing workflows. This enabled us to formalize a closed-loop system for an adaptive cruise control application, where a model-predictive controller was substituted with a neural network that had been thoroughly trained. Additionally, we were able to define and verify safety requirements for the system-controlled variables.

The proposed approach showed considerable improvements in verification efficiency, drastically reducing the time overhead compared to other leading techniques. Future work will focus on optimizing the interval arithmetic method, addressing the limitations related to overapproximation of results, and enhancing interval computation performance through GPU acceleration. Furthermore, we plan to extend this research by incorporating more detailed vehicle dynamics, including modeling behavior in PVS with a set of differential equations to account for more complex scenarios.

**Gergő Békési, Lilla Barancsik, Bálint Hartmann** (*Budapest University of Technology and Economics, Budapest, Hungary*)

## **DEEP NEURAL NETWORK STATE ESTIMATION WITH HYPERPARAMETER OPTIMIZATION.**

**Abstract:** *Machine learning methods are widely used to develop control components for cyber-physical systems, where control dynamics arise from subsystem integration. However, they may not always meet the reliability requirements for safe operation in critical environments. This paper presents an approach combining interval arithmetic and theorem-proving techniques to verify safety properties in closed-loop systems with neural network components. We apply this method to a model-predictive controller for autonomous driving, comparing the verification results with existing tools. Our findings show that interval arithmetic-based verification outperforms current methods, offering faster property proofs, and enabling formal modeling in higher-order logic for the controlled system.*

**Keywords:** *machine learning, control components, cyber-physical systems, safety properties, interval arithmetic, theorem-proving, autonomous driving, verification*

### **Introduction.**

Over the past decade, DSSE has gained significant attention in power system research due to the increasing role of distribution networks in energy production, especially with the decline of centralized generation. Distribution networks have become more important as they incorporate renewable energy sources and smart grid technologies, improving the efficiency, sustainability, and reliability of power grids.

The section highlights the role of DSSE in improving network monitoring and analysis in real-time. Distribution systems are critical to the sustainability and reliability of power grids, especially as they integrate renewable energy sources. The paper reviews several methods for DSSE, including traditional model-based algorithms like Weighted Least Squares (WLS), which require detailed knowledge of network parameters, and more advanced approaches such as machine learning and optimization techniques. Among the modern approaches, Deep Neural Networks (DNNs) have shown promise due to their ability to model complex, non-linear relationships in large datasets, though they require substantial computational resources.

The introduction also discusses various approaches to DSSE, categorizing them into model-based, forecasting-aided, and data-driven techniques. While traditional model-based methods are still widely used, the shift towards data-driven methods, particularly those using DNNs, is seen as a promising future direction due to their scalability and flexibility. The paper focuses on improving the accuracy of DNN-based DSSE using a hyperparameter optimization technique called Tree-structured Parzen Estimator (TPE), which is demonstrated on real-life low-voltage distribution networks in Hungary.

This approach marks a significant improvement over traditional methods like WLS by optimizing hyperparameters for better accuracy and computational performance. The section concludes by outlining the potential advantages of DNN-based DSSE with TPE in terms of accuracy, efficiency, and its applicability in real-world systems, especially for tasks such as renewable energy integration and real-time monitoring.

This part discusses the datasets used for the study, focusing on real-life low-voltage (LV) supply areas in Hungary. These networks were modeled to address challenges in obtaining network topology information, using data from Distribution System Operator (DSO) systems. The areas were selected based on their complexity and size, with each containing various circuits connected to an external grid element, a medium-voltage to low-voltage (MV/LV) transformer, and associated network components.

The four areas identified by numbers 18680, 44600, 44333, and 20667 are considered. Each area consists of different numbers of circuits, ranging from 2 to 10. The method used for modeling

LV networks adapts to data with inconsistent or uncertain lengths, utilizing available network data for better accuracy in simulating realistic scenarios.

In terms of input and output, the primary neural network applied in the study was a Fully Connected Neural Network (FCNN). The neural network used 15-minute interval data from both pseudomeasurements and weather data, such as temperature, sun altitude, radiation, wind speed, and precipitation. These data were used to improve the accuracy of the state estimation by accounting for external factors influencing consumption patterns. The data was also normalized and transformed into cyclical encodings to ensure proper learning performance.

The inputs to the network included power consumption, time data, and weather conditions, while the outputs were voltage amplitudes and angles. The data was divided into two sets: one for training and the other for validation. Training was performed on data from odd weeks, and validation was done using data from even weeks. This setup helped assess the accuracy and generalizability of the neural network model.

By applying this method, the study demonstrated that neural networks could effectively model and estimate power network states while integrating weather data and temporal factors, leading to improved accuracy compared to traditional estimation methods.

The authors introduce their approach for optimizing deep neural networks (DNNs) for Distribution System State Estimation (DSSE) tasks, specifically in the context of low-voltage (LV) networks. They begin by detailing the architecture of the fully connected neural network (FCNN) used in the study. FCNN was chosen for its simplicity, which allows for a clear comparison of how different hyperparameters affect network performance. The first layer of the FCNN computes the output by multiplying the input data with a weight matrix and adding a bias term, followed by non-linear activation using the ReLU function. This non-linearity is key for enabling the network to capture complex relationships in the data.

The paper also compares the FCNN to other competing architectures, such as Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) units, as well as Graph Neural Networks (GNNs). These alternatives are noted for their ability to process sequential data and model spatial relationships, respectively. For example, RNNs excel at capturing temporal dependencies, while GNNs are particularly effective in situations where data is structured as a graph, as in power grid networks. However, the study ultimately concludes that the FCNN with hyperparameter optimization outperforms these other models in the specific context of DSSE for low-voltage networks.

Training these neural networks involves optimizing a number of hyperparameters, and the authors use the Tree-structured Parzen Estimator (TPE) algorithm for this purpose. TPE is a Bayesian optimization method that models the hyperparameter space probabilistically, allowing it to more efficiently search for the best configuration compared to traditional random search or grid search methods. This approach allows the authors to achieve better results in terms of estimation accuracy and computational efficiency. Notably, TPE is shown to significantly outperform random search, and even the WLS method, in terms of reducing error and computational overhead.

By comparing the results of these different architectures and optimization strategies, the authors demonstrate the effectiveness of FCNN combined with TPE for DSSE tasks. Their findings emphasize the critical role of hyperparameter optimization in improving the performance of neural network-based state estimation methods for real-world low-voltage networks.

This part focuses on the experimental results of applying a deep neural network (DNN) for distribution system state estimation (DSSE) in low-voltage networks. The results demonstrate significant improvements in accuracy and efficiency compared to traditional methods such as the weighted least squares (WLS) method.

The DNN-based approach significantly reduces the relative error in voltage amplitude and angle estimations, achieving improvements ranging from 14% to 73%, depending on the network's size. This demonstrates the DNN's robustness and its ability to handle complex nonlinear relationships in real-world datasets. In addition to improving estimation accuracy, the method also provides substantial performance gains in terms of speed. Using GPU parallel computing, the DNN-

based model was able to complete the state estimation process in a fraction of the time required by the WLS method. In some cases, the speed-up was over 20,000 times faster.

The results also indicate that hyperparameter optimization (HPO) plays a crucial role in optimizing the DNN model. The study utilizes the Tree-structured Parzen Estimator (TPE) algorithm, which is a Bayesian HPO method. TPE proved to be more efficient than simpler optimization methods like random search, achieving faster convergence and more accurate results. Moreover, the study reveals that certain hyperparameters, such as batch size, learning rate, and hidden layer configuration, are correlated with the size of the distribution network, which influences the overall accuracy of the state estimation.

In conclusion, the DNN approach, supported by HPO, offers a more accurate and efficient solution for DSSE in low-voltage networks compared to traditional methods, making it a valuable tool for the integration of renewable energy sources and real-time monitoring of power systems.

The results of the hyperparameter optimization process demonstrate significant improvements in the accuracy and performance of the neural network. In most cases, the optimal values for hyperparameters such as batch size, learning rate, and hidden layer sizes are closely linked to the size of the distribution network. For larger areas, the HPO algorithm enables the use of larger batch sizes, speeding up the training process without sacrificing accuracy. This is particularly beneficial for more complex parts of the distribution system.

The analysis shows that the FCNN model optimized with HPO outperforms other models such as LSTM, GRU, and GNN in terms of accuracy and efficiency. The results also reveal that optimizing the hidden layer sizes has a more significant impact on larger networks, while other parameters like learning rate and step size optimize the training process more efficiently for larger networks. The proposed method achieves faster computation times and better estimations compared to traditional methods like the weighted least squares (WLS) approach.

In conclusion, the implementation of the Tree-structured Parzen Estimator (TPE) for hyperparameter optimization in FCNN-based DSSE significantly improves the estimation accuracy and efficiency of state estimation in low-voltage networks.

**Giampiero Bardella, Simone Franchini, Pierpaolo Pani, Stefano Ferraina** (*Sapienza University of Rome, Rome, Italy*)

## LATTICE PHYSICS APPROACHES FOR NEURAL NETWORKS.

**Abstract:** *Contemporary neuroscience has become an advanced field that integrates insights from various disciplines, leading to the development of innovative conceptual frameworks influenced primarily by physics and the study of complex systems. In this context, we recently proposed a mathematical model based on lattice field theory—an essential paradigm in theoretical particle physics—to describe the spatiotemporal interactions within neural networks. This brief overview outlines the fundamentals of our approach, aiming to be accessible to the broader neuroscience community. We position our methods in context, demonstrating how they can be directly linked to experimental parameters through familiar renormalization techniques. This summary introduces the essential concepts required for applying lattice physics to neural networks. These methods are particularly relevant in an age of rapid advancements in computational power, as they offer a pathway to connect observed neural dynamics with generative models grounded in physical principles.*

**Keywords:** *neuroscience, interdisciplinary, conceptual, physics, systems, lattice, field theory, neural spatiotemporal, interactions, networks, generative, models, computational, renormalization, dynamics*

### Introduction:

In recent years, neuroscience has made great strides in its efforts to understand how networks of neurons interact in both space and time, ultimately leading to cognition and behavior. Technological advancements have significantly improved our ability to record neural activity with high temporal resolution, allowing us to monitor the activity of hundreds of neurons simultaneously. Despite these advances, the data gathered has yet to yield clear answers about the fundamental mechanisms that govern neural systems, how neuronal connectivity relates to dynamics, and how complex behaviors emerge from these interactions.

The challenge in neuroscience today is akin to the situation in particle physics before the development of the Standard Model. Although there have been various models proposed to explain neural activity, the field still lacks a comprehensive, mechanistic understanding grounded in the deep physical principles that could unify these observations. The progress made in this area relies heavily on combining experimental data, computational models, and theoretical frameworks.

One such advancement has been the application of methods from theoretical particle physics, specifically quantum field theory (QFT), to understand systems of interacting binary variables like the spiking activity of neurons. These mathematical approaches help to translate complex neural interactions into terms that are grounded in universal laws of physics. The application of these methods reduces the gap between abstract theoretical models and real experimental data, offering new ways to link observed neural activity to models that are underpinned by physical principles.

In addition, the focus on these approaches is particularly relevant in an era marked by significant progress in computational power. With the help of these new computational tools, it is becoming increasingly possible to connect empirical observations of neural activity to generative models. These models are not only based on observed data but also follow the laws of physics, creating a new way to understand how complex neural systems function and how they relate to physical principles.

### Background:

In the field of computational neuroscience, there are currently two dominant perspectives: manifold modeling and circuit modeling. Manifold modeling posits that neural computations can be described by embedding the high-dimensional state space of neural dynamics into lower-dimensional manifolds. This approach suggests that understanding neural activity is about identifying the statistical relationships within the data without necessarily linking them to causal mechanisms. These methods, such as principal component analysis (PCA), reduce the data into lower-dimensional spaces that maximize variance, and they are effective for exploratory analysis and predicting experimental

variables. However, they often provide results that lack interpretability in the context of general principles and can lead to contradictory conclusions.

On the other hand, circuit modeling relies on the assumption that the connectivity between neural units forms the foundational mechanism for brain function. This approach integrates realistic biophysical parameters like membrane potentials and cell types, allowing for the prediction of neural activity through a set of specific connections. While successful in modeling certain neural behaviors, such as stimulus selectivity and neural dynamics, circuit models are often criticized for being highly parameter-dependent, making it difficult to scale them to larger networks. These models also face challenges in explaining mixed selectivity and other phenomena that involve multiple, interacting variables.

Neither approach fully integrates with general physical principles, which creates a gap in our understanding. One promising method that attempts to bridge this gap is neural field theory (NFT). NFT models large-scale neural activity using differential equations and considers anatomical and physiological details, offering a more comprehensive way to describe brain dynamics. However, like other approaches, it faces challenges in linking the various scales at which neural processes occur.

In the realm of physics, the most successful models have applied statistical physics principles to neural networks. For example, the Amari-Hopfield model, a prominent framework in computational neuroscience, draws parallels between the energy function in neural systems and the energy in spin systems from magnetic physics. This connection allows for the use of stochastic dynamics and bifurcation theory to describe neural behavior. Yet, these models often struggle with exact solutions, particularly when dealing with large systems, and the challenge of linking them to empirical data remains.

In response to these limitations, some researchers have argued that quantum field theory (QFT) could provide a more robust framework for understanding neural systems. By treating neural interactions as systems of discrete binary variables, much like the interactions of particles in physics, this approach could offer a more direct connection between theoretical models and experimental data.

### **Preliminaries:**

To dive deeper into the framework we propose, let's briefly discuss the Lagrangian description of a dynamic system, a key component of our model. The Lagrangian approach, while less well-known outside of physics, offers an essential perspective. Unlike other methods, such as Newtonian mechanics, the Lagrangian framework focuses on energy, generalized coordinates, symmetries, and conservation laws. This method is particularly useful for complex systems with many degrees of freedom, as it unifies a wide range of physical phenomena, from classical mechanics to electromagnetism and general relativity.

At its core, the Lagrangian represents the difference between kinetic and potential energy. The action, which is the integral of the Lagrangian over time, encapsulates the dynamics of the system. The principle of least action, one of the fundamental laws of physics, asserts that the path a system takes between two states minimizes the action. This idea, which parallels Newton's laws of motion, is a powerful tool for describing the evolution of physical systems.

In the context of neuroscience, most foundational models do not yet incorporate the principle of least action. While some indirect evidence exists supporting the idea, a systematic Lagrangian approach to neural dynamics has not yet been fully explored. Reimagining neural circuits using this framework would allow neuroscience to engage more directly with the broader theoretical landscape of physics, borrowing analytical techniques and numerical methods from these fields. If physics consistently operates under this principle, it seems only logical to extend this approach to neural circuits as well.

### **Fundamentals:**

Our proposed model introduces a shift from continuous spacetime to a discrete lattice, converting fields into variables defined at specific lattice sites at discrete time steps. This approach is grounded in lattice field theory (LFT), a standard computational technique in quantum field theories (QFT), especially useful when analytical methods are not feasible. The lattice formalism guarantees



the preservation of symmetries and conservation laws, even when dealing with a discrete approximation of continuous space.

We specifically focus on binary activity within neural networks. Neurons are represented as binary variables—either 0 (inactive) or 1 (active)—which reflects their spiking behavior. This simplification, though basic, captures the essence of neural interactions in a compact form. Each neuron in the network is mapped to a lattice vertex, and the temporal evolution of the neural network is discretized over time. This allows us to model the spiking activity of neurons in a lattice structure, where each site represents a neuron, and the interactions between neurons are represented as interactions between the corresponding lattice points.

The time intervals between computational steps are chosen to be the smallest possible value, corresponding to the duration of a spike. In this framework, the minimum relevant timescale is tied to the refractory period of the neurons, which typically lasts about 1 millisecond. The activity of the neurons is then captured in a binary array, known as the "kernel," where each row represents a neuron, and each column represents a time step. This kernel contains the temporal sequence of spikes recorded for a network of neurons during an experiment.

The spatial correlations between neurons are captured by a matrix that quantifies pairwise correlations over the entire time period, while the temporal correlations represent how neurons' spiking patterns relate to each other across time. These matrices, known as the hypermatrix, allow us to model the collective behavior of neural activity as a set of spatiotemporal patterns, making it easier to analyze and interpret the interactions within the network.

#### **Modeling Neural Networks with the Action:**

In our framework, the evolution of neural dynamics is modeled as a process influenced by past states, akin to a quantum system with memory. This approach, while initially seeming unconventional, simplifies the problem remarkably. The classical (non-quantum) evolution of a system can always be derived from the quantum version, meaning that our model operates within a broad and flexible theoretical framework.

We treat the neural network as a discrete system of interacting binary fields, analogous to qubits in quantum mechanics. This simplification allows us to apply lattice-based statistical mechanics to describe neural interactions, enabling the use of powerful methods from quantum field theory (QFT) to solve the system. The challenge of determining the temporal evolution of such a system can be tackled through statistical mechanics techniques applied to the lattice, providing us with an efficient way to study complex neural networks.

The action function in this model represents the system's behavior over time. Just as in particle physics, the action function defines how the system evolves by balancing kinetic and potential energies. In the case of neural networks, the action function encodes the collective behavior of neurons over time, capturing both the spatial and temporal interactions between them. By computing the action of the system, we can understand how the network of neurons interacts, how information is transmitted, and how these interactions change over time.

#### **Renormalization and Effective Theories:**

One of the key challenges in relating theory to experimental data is the concept of renormalization, which allows us to simplify complex systems by focusing on relevant scales. In neural networks, this process involves simplifying the system's behavior at different spatial or temporal scales while retaining the essential dynamics. Renormalization allows us to connect microscopic details (such as individual neurons or synapses) with macroscopic observations (such as large-scale brain activity) without needing to model every detail of the system.

In our framework, renormalization is applied by decimating the lattice, essentially averaging over smaller-scale details to produce a simplified model that still captures the essential features of the neural network. This decimation process is particularly useful when dealing with experimental data that cannot capture every individual neuron's behavior but can provide insights into the collective activity of larger groups of neurons. By applying renormalization, we can connect the data recorded from experiments (such as electrode arrays) with the theoretical model, providing a more accurate and comprehensive understanding of neural interactions.

For example, when analyzing data from a Utah array, which samples activity from a grid of electrodes, the neural dynamics can be modeled as a decimated lattice. This method allows us to treat the entire minicolumn as a single unit, simplifying the modeling process while still capturing the relevant dynamics. The use of renormalization by decimation ensures that we focus on the most significant interactions in the system, leading to a more manageable and interpretable model.

#### **Learning Neural Interactions from Data:**

One of the main goals of this approach is to learn about the underlying interactions within a neural network from experimental data. By measuring quantities like average spiking activity and correlations between neurons, we aim to infer the neural interactions that govern the observed dynamics. This inverse problem, where we aim to infer the model parameters from experimental data, is central to understanding how neural circuits function.

To solve this problem, we use statistical models like maximum entropy models, which provide a way to derive the parameters of the network from the observed data. These models make minimal assumptions about the system, ensuring that the derived parameters capture the true underlying interactions. By applying these models to neural data, we can learn about the connectivity and dynamics of the network.

One important aspect of our model is its ability to capture both the spatial and temporal aspects of neural interactions. Unlike traditional maximum entropy models, which focus on the spatial correlations of neural activity, our model includes a kinetic term that captures the temporal evolution of the system. This allows us to study how neurons influence each other not only in space but also over time, providing a more complete description of neural dynamics.

#### **Applications and Future Directions:**

The framework we propose offers exciting possibilities for studying neural systems, both at the level of individual neurons and across large networks. By integrating data from multiple sources, including electrophysiological recordings, imaging, and behavioral data, we can build more comprehensive models of brain activity. These models can be used to simulate neural dynamics, predict neural responses to different stimuli, and explore how changes in connectivity or dynamics affect brain function.

The approach also opens up new possibilities for generative modeling of neural activity. By using the principles of lattice field theory, we can generate synthetic neural data that mimics the observed dynamics of real neural networks. This could be useful for testing hypotheses about neural function and for developing new techniques for analyzing neural data. Additionally, the ability to model neural networks within a physical framework could lead to a deeper understanding of the brain's computational principles, potentially unlocking new insights into cognition, learning, and behavior.

The flexibility of our approach makes it adaptable to a wide range of neural systems, from small circuits of neurons to large-scale brain networks. By refining the renormalization techniques and incorporating more sophisticated models of neural activity, we can continue to improve our understanding of how the brain works, both in health and disease.

#### **LFT and Biohybrids Networks:**

The recent progress in integrating biological brain tissue with bio-inspired electronics has opened exciting new avenues for understanding and utilizing neural systems. These advancements have led to the development of brain organoids and neuromorphic systems, which represent significant steps forward in both biological research and technology. Brain organoids—miniaturized, simplified versions of the brain—allow researchers to model neural networks in a laboratory setting, while neuromorphic systems are electronic devices designed to simulate the brain's computational processes.

The integration of biological neurons with artificial systems, also known as biohybrids, is a particularly promising area of research. The potential to combine the flexibility and learning capabilities of biological networks with the precision and efficiency of artificial systems could lead to novel applications in both neuroscience and technology.

Our framework, based on lattice field theory (LFT), could play a critical role in the development of these biohybrid systems. By using LFT to model neural dynamics in these systems, we can more accurately simulate how artificial and biological components interact. This could lead to more efficient designs for neuromorphic chips, which, like the human brain, process information only when events (such as spikes) occur. Such chips could be energy-efficient while maintaining the adaptability and learning capabilities of biological neural networks. LFT offers a powerful tool to describe the dynamics of these biohybrids, as it allows us to model their activity in terms of discrete units (binary states) and map these onto experimental data.

Moreover, by applying LFT in this context, we can explore how artificial systems, such as neuromorphic chips, can be fine-tuned based on the activity of recorded biological neurons during both the design and testing phases. This creates the potential for more efficient and biologically plausible artificial neural networks, which could have applications ranging from brain-machine interfaces to advanced AI systems.

#### **Limitations of the Study:**

While our framework offers numerous advantages, there are several open challenges that need to be addressed before it can be fully implemented. One of the main challenges lies in linking the effective theory, which describes the neural network at larger scales, with the microscopic theory that focuses on individual neurons. This requires the use of renormalization procedures, which are still in the early stages of development. A more detailed understanding of how to apply these procedures across different scales will be crucial for advancing the framework.

Currently, we employ approximations, such as the two-body truncation, to simplify the system. This approach is valid when covariances between neurons are small, but it may not hold in all cases. Additional approximations, such as stationary conditions truncation, are applied when synaptic connections remain relatively constant over time. These approximations are valid for most neurophysiology experiments, where the scale of changes in brain activity is manageable. However, in experiments with highly dynamic synaptic properties or networks with rapidly changing connectivity, these simplifications may not be sufficient.

Another challenge is mapping the observed dynamics of the neural system to its underlying anatomical structure. While we assume that the anatomical structure of neurons and synapses is reflected in the observed dynamics, the exact correspondence between the coupling parameters and the physical structures is not trivial. In some cases, it may be difficult to determine these structures purely from observable data, especially when there is insufficient precision in the measurements.

Finally, like many models in computational neuroscience, our framework faces the challenges associated with inverse problems—problems where the goal is to infer the underlying parameters of a system from limited or noisy data. The large number of parameters in the model, combined with the relatively small amount of experimental data, makes this task particularly challenging. However, advances in computational methods and machine learning techniques may help address these issues in the future, enabling more accurate inference from experimental data.

#### **Theoretical Insights:**

At the heart of our model lies the concept of the action function, which is used to describe the dynamics of neural networks. The action is the core of the system's evolution and is formulated in terms of the Lagrangian, which is the difference between the system's kinetic and potential energy. In the case of neural networks, the action incorporates both the interactions between neurons and their temporal evolution, providing a unified framework for understanding the behavior of large-scale neural systems.

To compute the action, we utilize a statistical field theory approach, where the action is expanded as a Taylor series. This allows us to describe the interactions in a neural network at multiple levels of complexity, from one-vertex interactions (simple neuron activity) to higher-order interactions (complex patterns of neural firing). The mathematical formalism enables us to analyze how the system's state evolves over time, providing insights into how different parts of the network influence one another and how these interactions shape brain function.

By focusing on binary variables for neuron activity (spike or no spike), we simplify the description of neural interactions, making the model more computationally tractable. This simplification does not undermine the complexity of the system but rather allows us to capture its essential dynamics in a manageable form. As we continue to refine the model, we can extend it to account for more complex neurophysiological signals, such as local field potentials (LFPs) or multi-unit activity (MUA), providing a more comprehensive description of neural function.

#### **Renormalization and Multiscale Analysis:**

One of the most powerful features of our framework is its ability to integrate renormalization techniques, which are essential for scaling up the model from individual neurons to large-scale networks. Renormalization allows us to "zoom in" or "zoom out" on different parts of the system, focusing on the relevant scales while ignoring the less important details. This process is particularly useful when dealing with experimental data, which often cannot capture every neuron in a network but can provide valuable insights into the collective behavior of the network as a whole.

In practice, we apply a renormalization-by-decimation approach, where we simplify the system by integrating out the small-scale details (such as individual neurons) and focusing on the larger-scale patterns (such as neural minicolumns). This method preserves the essential dynamics of the system while making the model more computationally feasible.

For example, when analyzing data from a Utah array, which records neural activity from a grid of electrodes, we can model the neural dynamics as a decimated lattice, where each electrode corresponds to a larger group of neurons (a minicolumn). This decimation process allows us to focus on the most relevant interactions within the neural network while simplifying the overall model. By applying these techniques, we can make accurate predictions about neural activity across different spatial and temporal scales, providing valuable insights into brain function.

#### **Learning Neural Interactions from Data:**

Understanding neural network interactions involves determining how the activity of neurons is organized and how these interactions contribute to overall brain function. To achieve this, we must infer the underlying neural connections from experimental data. This is known as the inverse problem, where the goal is to extract model parameters (such as coupling strengths between neurons) based on observed neural activity. This process is crucial for advancing our understanding of how neural networks operate and for developing accurate models of brain activity.

One approach to solving the inverse problem is through the use of maximum entropy models. These models maximize the entropy, or uncertainty, of a system subject to certain constraints, ensuring that the inferred model makes the fewest assumptions about the network's behavior. This approach has been applied in neuroscience to infer the spatial couplings between neurons, but it often assumes that temporal dynamics are not explicitly modeled. While these models can capture basic interactions, they do not account for the temporal evolution of the system, which is a critical aspect of neural dynamics.

Our framework addresses this limitation by incorporating a kinetic term in the model. This term explicitly accounts for the temporal evolution of the system, allowing the model to capture both the spatial and temporal aspects of neural interactions. By doing so, we can infer not only how neurons are connected but also how they influence each other over time. This extended model can be applied to various types of neural networks, whether they exhibit asynchronous activity, oscillations, or other complex dynamic behaviors.

For example, the model can be used to simulate neural time series, estimate the coupling parameters from experimental data, and compare the results with theoretical predictions. It can also be used to study different dynamical regimes of the system, such as oscillatory or chaotic dynamics, by analyzing how the coupling parameters change under different conditions. This approach allows us to gain insights into how neural circuits operate, how their dynamics are influenced by external factors, and how these dynamics relate to the brain's functional behavior.

#### **Applications and Future Directions:**

The model we present has a broad range of applications, both in basic neuroscience research and in more applied fields, such as neuroprosthetics and brain-machine interfaces. One promising

area is the study of large-scale brain networks. By using the principles of lattice field theory (LFT), we can model these networks at multiple levels of organization, from individual neurons to whole brain regions. This multiscale approach enables a more comprehensive understanding of how different brain regions communicate and cooperate to produce complex cognitive functions.

Additionally, our framework could play a key role in the development of generative models for neural activity. These models would allow researchers to simulate the neural activity of large networks based on the observed dynamics of real systems. By comparing these simulations with experimental data, we can test hypotheses about neural function and improve our understanding of how the brain processes information. Furthermore, these generative models could be used to develop new tools for analyzing neural data, providing insights into the mechanisms underlying cognition, learning, and behavior.

In the realm of artificial intelligence, our approach could be used to improve the design of neuromorphic systems—artificial neural networks that mimic the structure and function of biological brains. By applying LFT to neuromorphic circuits, we could optimize their design to better replicate the dynamics of real neural systems. This could lead to more efficient and adaptable artificial systems, with applications in areas such as robotics, machine learning, and brain-computer interfaces.

Another exciting direction is the integration of our model with biological systems, such as brain organoids and biohybrids. By using LFT to model the neural dynamics in these systems, we can explore how artificial and biological components interact. This could pave the way for the development of hybrid systems that combine the strengths of both biological and artificial networks, offering new opportunities for both neuroscience research and the creation of bio-inspired technologies.

#### **Conclusion:**

The framework we have developed represents a powerful and versatile tool for studying neural dynamics. By applying lattice field theory (LFT) to neural networks, we can gain new insights into the complex interactions that underlie brain function. Our approach allows us to model neural systems at multiple scales, from individual neurons to large-scale brain networks, and provides a more unified and physically grounded understanding of neural dynamics.

While there are still challenges to overcome—such as refining renormalization techniques and addressing the limitations of inverse problems—the potential of this framework is vast. It opens up new possibilities for both basic neuroscience and applied fields, offering a path forward for more accurate models of brain activity, more effective brain-machine interfaces, and a deeper understanding of cognition and behavior.

As computational power continues to advance, we expect that these models will become even more sophisticated and capable of addressing the growing complexity of neural systems. By linking experimental data to physical principles, our approach provides a novel perspective on how the brain operates and how it can be modeled, offering exciting prospects for the future of neuroscience and neuroengineering.

**Lilli Frison, Simon Gölzhäuser, Moritz Bitterling, Wolfgang Kramer** (*Fraunhofer Institute for Solar Energy Systems ISE, Freiburg, Germany*)

## **ARTIFICIAL NEURAL NETWORKS FOR OPTIMIZING DISTRICT HEATING NETWORKS.**

**Abstract:** *Accurate forecasting of heat demand is crucial for optimizing the energy efficiency of district heating networks (DHNs), which are subject to challenges such as fluctuating weather, unpredictable user behavior, and varying energy supply. This study assesses the performance of Artificial Neural Networks (ANNs), including advanced models like Long Short-Term Memory (LSTM) Networks, Convolutional Neural Networks (CNNs), and the Temporal Fusion Transformer (TFT), compared to the traditional SAR-IMAX statistical model. The evaluation focuses on each model's capacity to predict varying heat demand profiles and deliver interpretable forecasts with optimization techniques, particularly highlighting the importance of clear confidence intervals. Using a year's worth of data from the Stiftung Liebenau DHN, which integrates multiple energy sources such as Combined Heat and Power (CHP), biomass, and natural gas, along with diverse heat sinks like residential buildings and greenhouses, the study finds that while the CNN model achieves the lowest Mean Absolute Percentage Error (MAPE)—27% for both summer and winter, and 17% for winter alone—accuracy is notably influenced by data volatility and irregularities. Despite this, the models successfully capture general trends, though predicting demand peaks and fluctuations remains challenging. An economic assessment reveals that these predictive methods significantly improve energy efficiency and provide economic advantages due to their low initial investment requirements.*

**Keywords:** *heat demand, forecasting, district heating, artificial neural networks, energy efficiency, prediction accuracy, economic analysis, confidence intervals*

### **Introduction:**

The accurate forecasting of heat demand is crucial for optimizing the operation of district heating networks (DHNs), which face challenges such as variable weather, fluctuating user behavior, and energy availability. To address this, the study examines the effectiveness of various artificial neural network (ANN) models, including Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), and the Temporal Fusion Transformer (TFT), and compares them to the traditional statistical SAR-IMAX model. These models are evaluated for their ability to predict diverse heat demand profiles and produce interpretable forecasts with optimization strategies, especially focusing on understandable confidence intervals. Using data from Stiftung Liebenau DHN, which incorporates multiple energy sources such as Combined Heat and Power (CHP), biomass, and natural gas, and various heat sinks including residential buildings and greenhouses, the study finds that the CNN model, while achieving the lowest Mean Absolute Percentage Error (MAPE) of 27% for both summer and winter and 17% for winter alone, is significantly influenced by data volatility and irregularities. However, the models successfully capture overall trends, though predicting demand peaks and fluctuations proves challenging. An economic analysis indicates that applying these forecasting methods can significantly improve energy efficiency and provide economic benefits due to low investment costs.

The study aims to enhance the forecasting of heat demand for district heating networks (DHNs) by applying advanced Artificial Neural Network (ANN) models. The models assessed include Long Short-Term Memory (LSTM) Networks, Convolutional Neural Networks (CNN), and Temporal Fusion Transformers (TFT). These models were compared against the statistical SAR-IMAX model to evaluate their effectiveness in predicting heat demand and providing reliable, interpretable forecasts. The key focus of the research is on the models' ability to offer optimization strategies with clear confidence intervals. To test the models, data from Stiftung Liebenau DHN, collected over the course of a year, was used. This data set included multiple energy sources, such as Combined Heat and Power (CHP), biomass, and natural gas, and encompassed a wide range of heat sinks, including residential buildings, greenhouses, and other facilities.

Despite the CNN model delivering the best performance, with a Mean Absolute Percentage Error (MAPE) of 27% for both summer and winter and 17% for winter alone, the accuracy of the models was influenced by the volatility and irregularities present in the data. While the models could

capture general trends, they faced difficulties in predicting peak demands and fluctuations. Nevertheless, an economic analysis revealed that the application of these predictive models could significantly enhance energy efficiency. Moreover, their relatively low investment costs mean that they offer substantial economic benefits.

The study highlights the importance of accurate forecasting in optimizing the operation of district heating networks (DHNs), particularly in the context of 4th generation systems. These networks, which deliver heat from central sources to various buildings, are essential for urban energy infrastructures. The transition towards 4th generation systems involves integrating renewable energy sources, improving the efficiency of heat distribution, and adapting to varying heat demands. One of the key challenges of such systems is the integration of decentralized and fluctuating renewable energy sources, as well as the need for bidirectional heat transfer with buildings, efficient energy storage, and low-temperature operations.

Given the complexity of these systems and the diverse consumers they serve, it is crucial to develop advanced predictive techniques that can process large volumes of data and account for numerous variables influencing heat demand. By using predictive models, energy stakeholders can better anticipate heating requirements, leading to more efficient energy distribution, reduced operational costs, and greater energy sustainability. Traditional statistical methods for heat load forecasting have evolved into more sophisticated Artificial Neural Networks (ANNs), which offer improved accuracy and adaptability in modeling complex, nonlinear relationships in heating demand patterns. The growing availability of data from smart meters and IoT devices has significantly enhanced the capabilities of ANNs in this domain. These advanced models can process real-time data and respond more dynamically to changing conditions, ultimately contributing to more efficient and sustainable energy systems.

The paper explores various machine learning methods for time series forecasting, with a particular focus on Artificial Neural Networks (ANNs) and Deep Neural Networks (DNNs). These techniques are increasingly being used to predict thermal load for district heating networks (DHNs), which involves forecasting both individual heat consumers and the overall demand of the entire network. The study highlights how the growing availability of data, as well as advances in computational power, have made traditional forecasting methods less effective compared to AI-based approaches.

Neural network architectures that include recurrent components, such as Long Short-Term Memory (LSTM) networks, are often employed for handling sequential data, which is a key feature of heat demand patterns. Other advanced architectures, such as Temporal Convolutional Networks (TCNs), are also utilized for improving forecasting accuracy. Moreover, hybrid approaches combining DNNs with traditional time series analysis methods are emerging as powerful solutions for improving prediction accuracy.

The research addresses the practical challenges of applying Artificial Neural Networks (ANNs) in real-world district heating network (DHN) applications. While ANNs show great promise in predicting heat demand, their real-world implementation faces difficulties due to the diversity of consumers within a DHN and the insufficient availability of high-quality data. To improve the accuracy of predictions, the study emphasizes the need for uncertainty estimates, model interpretability, and the integration of advanced machine learning techniques.

The study compares two standard ANN architectures: Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. It also explores two different approaches for multi-step predictions—single-shot and autoregressive prediction. These approaches are evaluated to determine which is more suitable for forecasting heat demand across different consumers, including residential buildings, offices, and greenhouses. Additionally, the Temporal Fusion Transformer (TFT) network is considered for comparison, as it is specifically designed for time series forecasting. The research further explores how these models can be integrated into control algorithms to enhance operational efficiency.

The paper uses the Stiftung Liebenau DHN, a small local community DHN in Meckenbeuren, Germany, as a case study. This DHN features multiple heat sources, including Combined Heat and

Power (CHP), natural gas, and biomass, as well as various heat sinks such as residential buildings, workshops, greenhouses, and medical facilities. Through this case study, the paper not only evaluates the technical aspects of implementing ANNs for heat demand forecasting but also addresses the real-world challenges of data collection, processing, and integration into existing DHN systems. The results suggest that ANNs can significantly improve heat demand forecasting accuracy, but practical challenges such as data accessibility and the integration of predictive models into control systems remain areas for further development.

The paper examines the various artificial neural network (ANN) models employed for forecasting heat demand in district heating networks (DHNs). The study assesses the effectiveness of these models, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and the Temporal Fusion Transformer (TFT), for accurately predicting heat load profiles and providing reliable forecasts. The models are compared to the statistical SAR-IMAX model to evaluate their relative performance in real-world applications.

Using a year's worth of data from Stiftung Liebenau DHN, which encompasses multiple energy sources like Combined Heat and Power (CHP), biomass, and natural gas, the study demonstrates the potential of these AI-based methods. The data also includes diverse heat sinks, such as residential buildings and greenhouses, which add complexity to the heat demand forecasting process. The CNN model, despite achieving the lowest Mean Absolute Percentage Error (MAPE) of 27% for both summer and winter, and 17% for winter alone, faces challenges in prediction accuracy due to the volatility and irregularity of the data. Nonetheless, the models successfully capture the overall heat demand trend, though they struggle to predict peaks and fluctuations in demand. The economic analysis further suggests that, despite the challenges, applying these predictive models could significantly improve energy efficiency while offering economic benefits due to their relatively low investment costs.

The paper focuses on the practical implementation and evaluation of Artificial Neural Networks (ANNs) for forecasting heat demand in district heating networks (DHNs). The study specifically explores several ANN architectures, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and the Temporal Fusion Transformer (TFT). These models are tested against the traditional SAR-IMAX statistical model to assess their relative effectiveness in predicting heat load and providing actionable forecasts.

The research uses data from the Stiftung Liebenau DHN, a local district heating network located in Meckenbeuren, Germany. This data spans one year and incorporates diverse energy sources, such as Combined Heat and Power (CHP), biomass, and natural gas, as well as a variety of heat sinks, including residential buildings and greenhouses. These diverse components of the DHN provide a complex landscape for heat demand forecasting, making the role of AI models particularly important. The study reveals that while the CNN model demonstrates the lowest Mean Absolute Percentage Error (MAPE) — achieving 27% for both summer and winter, and 17% for winter alone — prediction accuracy is notably affected by data volatility and irregularity.

Despite these challenges, the AI models effectively capture the general heat demand trends, though they struggle to predict sharp demand peaks and fluctuations. The paper further explores the economic implications of using these predictive models, concluding that they can significantly enhance energy efficiency within DHNs, offering potential economic benefits due to their low initial investment costs. The results underscore the promise of AI in optimizing heat demand forecasting, despite the complexities of real-world data.

The study highlights the significance of accurate forecasting for the optimal operation of district heating networks (DHNs), particularly as these systems evolve towards more sustainable and flexible models. Traditional DHNs, which have been central to urban heating infrastructures, are now transitioning into 4th generation systems that incorporate renewable energy sources, lower temperature operations, and advanced control strategies. These modern systems face unique challenges, including integrating decentralized and variable renewable energy inputs, managing bidirectional heat transfer with buildings, and customizing heat supply to meet dynamic demands.



As the complexity of these networks grows, advanced predictive techniques become essential. Such techniques must handle large volumes of diverse data and account for the various factors influencing heat demand, including weather conditions, user behavior, and the availability of energy sources. By utilizing predictive models, energy providers can anticipate heating needs with greater accuracy, optimizing energy distribution, reducing operational costs, and achieving long-term energy efficiency goals. Predictive modeling helps balance supply and demand, manage peak loads, and integrate renewable energy more effectively, making it an essential tool for modern DHNs. The adoption of Artificial Neural Networks (ANNs) has emerged as a significant advancement, as these models can capture complex, nonlinear relationships within heat demand data, something that traditional statistical models struggle with. With the increasing availability of real-time data from smart meters and IoT devices, ANNs can process this data to improve the responsiveness and adaptability of DHNs, ultimately contributing to more efficient and sustainable heating systems.

The research delves into the application of Artificial Neural Networks (ANNs) for heat demand forecasting in district heating networks (DHNs). It evaluates several ANN architectures, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Temporal Fusion Transformers (TFTs), with the goal of enhancing the accuracy of heat load predictions. These models are compared against the traditional SAR-IMAX statistical model to assess their effectiveness in real-world applications.

Data from the Stiftung Liebenau DHN, which spans a year and includes a mix of energy sources like Combined Heat and Power (CHP), biomass, and natural gas, as well as various heat sinks such as residential buildings and greenhouses, is used for testing. The results show that while the CNN model achieves the lowest Mean Absolute Percentage Error (MAPE)—27% for both summer and winter, and 17% for winter alone—its performance is still affected by the volatility and irregularity in the data. However, despite this, the models successfully capture the general trend in heat demand. The study also finds that predicting demand peaks and fluctuations remains a challenge. Additionally, an economic analysis indicates that, despite the issues with data volatility, the use of these predictive models can significantly enhance energy efficiency and provide economic benefits due to their relatively low investment costs.

The study investigates the practical application of Artificial Neural Networks (ANNs) for forecasting heat demand in district heating networks (DHNs), with a focus on improving operational efficiency. The research compares multiple ANN architectures, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Temporal Fusion Transformers (TFTs), to traditional forecasting methods like the SAR-IMAX model. The goal is to assess how well these models predict heat demand and provide reliable forecasts that can aid in the optimization of DHN operations.

Using a year's worth of data from Stiftung Liebenau DHN, which integrates various energy sources such as Combined Heat and Power (CHP), biomass, and natural gas, as well as diverse heat sinks including residential buildings, workshops, and greenhouses, the study evaluates the performance of these AI-based models. Despite the CNN model achieving the lowest Mean Absolute Percentage Error (MAPE) of 27% for both summer and winter, and 17% for winter alone, the accuracy is still significantly affected by the irregularities and volatility present in the data. Although the models effectively capture overall heat demand trends, predicting fluctuations and demand peaks proves to be a more challenging task. However, the paper underscores that the use of these predictive models can still lead to substantial improvements in energy efficiency, offering clear economic benefits due to their relatively low initial investment costs. The research highlights the promise of ANNs for optimizing heat demand forecasting, even as challenges related to data volatility and prediction accuracy remain.

The paper explores the application of Artificial Neural Networks (ANNs) in forecasting heat demand within district heating networks (DHNs), aiming to improve operational performance and energy efficiency. It specifically evaluates several types of ANN architectures, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Temporal Fusion Transformers (TFTs), in comparison with the traditional SAR-IMAX statistical model. The primary

objective is to determine how well these models can predict heat demand and provide accurate, interpretable forecasts for optimal DHN management.

The study uses data collected over the course of one year from Stiftung Liebenau DHN, which involves various energy sources like Combined Heat and Power (CHP), biomass, and natural gas, as well as multiple heat sinks such as residential buildings, greenhouses, and workshops. The results indicate that while the CNN model achieves the lowest Mean Absolute Percentage Error (MAPE)—27% for both summer and winter, and 17% for winter alone—the accuracy of predictions is notably influenced by the volatility and irregularities in the data. Nevertheless, the models manage to capture the general trends in heat demand, even though predicting demand fluctuations and peak periods proves challenging.

Despite these challenges, the research concludes that the implementation of these predictive models can significantly enhance energy efficiency. The economic analysis suggests that these models offer substantial benefits, particularly due to their relatively low initial investment costs. While the models are not flawless, their ability to improve energy forecasting makes them a promising tool for the optimization of DHN operations.

## **Conclusions**

The study demonstrates the potential of Artificial Neural Networks (ANNs) in improving heat demand forecasting for district heating networks (DHNs). The application of advanced ANN models, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Temporal Fusion Transformers (TFTs), offers significant advantages over traditional statistical methods like the SAR-IMAX model, particularly in terms of capturing complex, nonlinear relationships within the heat demand data.

Despite challenges such as data volatility and irregularity, the models successfully forecast general trends in heat demand, although predicting demand peaks and fluctuations remains a challenge. Among the ANN models, the CNN architecture showed the best performance, achieving the lowest Mean Absolute Percentage Error (MAPE), particularly in the summer and winter seasons. However, the overall accuracy of these models is influenced by the inherent variability of the data.

From an economic perspective, the study finds that the application of these predictive models can result in substantial improvements in energy efficiency. The relatively low investment costs associated with implementing these models offer clear economic benefits, especially in terms of optimizing energy distribution and reducing operational costs. While further refinement is needed to enhance their predictive accuracy, these models hold significant promise for the future of DHN operation, contributing to more sustainable and resilient energy systems.

In conclusion, the integration of ANNs into DHN management has the potential to revolutionize heat demand forecasting, providing not only operational advantages but also economic and environmental benefits. The continued development and refinement of these models will play a crucial role in achieving the broader goals of energy efficiency and sustainability in district heating systems.

**Lukasz Sobolewski** (*University of Zielona Góra, Zielona Góra, Poland*)

## **QUALITY EVALUATION OF POLISH TIMESCALE FORECASTING USING GMDH NEURAL NETWORK.**

**Abstract:** *The article discusses the results of evaluating the effectiveness of a forecasting system that employs a Group Method of Data Handling (GMDH) neural network. This system was used to forecast the Polish Timescale (UTC(PL)) based on hydrogen maser (HM) technology. It highlights how the use of this forecasting system for implementing the UTC(PL) national timescale can yield a quality comparable to that of the best global timescales. The study presents the forecasting outcomes for UTC(PL) based on two prepared time series (TS1 and TS2), with a comparison to the UTC - UTC(k) values for the same scale. Excellent forecasting performance was achieved in cases 5 and 6, for both TS1 and TS2, as demonstrated by the forecast quality metrics. These findings indicate that the proposed system ensures the UTC(PL) accuracy on par with the top international timescales. The results are expected to encourage National Metrology Institutes (NMIs) lacking caesium fountain capabilities to adopt the UTC(k) steering system.*

**Keywords:** *forecasting, GMDH, neural network, UTC(PL), hydrogen maser, timescale, accuracy, National Metrology Institutes*

### **Introduction:**

The article discusses the effectiveness of a forecasting system for the Polish Timescale (UTC(PL)), using the Group Method of Data Handling (GMDH) neural network and hydrogen maser (HM) technology. The research aims to show that applying this forecasting system to UTC(PL) can achieve timescale quality comparable to the best timescales worldwide. Forecasting results are presented for two time series, TS1 and TS2, and compared to UTC - UTC(k) values. The results demonstrate excellent forecasting quality, with cases 5 and 6 achieving very good results, as confirmed by forecast quality measures. The research concludes that the developed system ensures UTC(PL) accuracy on par with the top global timescales, supporting its adoption by National Metrology Institutes (NMIs) lacking caesium fountains.

The research presented in the article focuses on the effectiveness of a forecasting system for the Polish Timescale (UTC(PL)), which is based on a GMDH (Group Method of Data Handling) neural network. This system, designed for the Polish national timescale, is implemented with a hydrogen maser (HM). The primary goal of the study is to demonstrate that the use of this forecasting system can ensure that the quality of the Polish Timescale is comparable to that of the best global timescales.

The forecasting system operates on two specially prepared time series, TS1 and TS2. These series, along with the forecasted values for the Polish Timescale, are compared with the UTC - UTC(k) values for this scale. The analysis confirms the high quality of the forecasts, particularly for cases 5 and 6, where the system showed a high degree of accuracy. This is evident from the forecast quality metrics and the minimal discrepancies observed between the predicted and actual values. These findings suggest that the developed forecasting system allows for the accurate control and correction of the UTC(PL) national timescale, placing it among the best-performing timescales globally.

The research is particularly significant for National Metrology Institutes (NMIs) that lack caesium fountains, as it shows that adopting the UTC(k) steering system based on this neural network approach could allow them to achieve similar high-quality results at a lower cost.

The second key factor influencing the forecasting quality is the quality of the prepared input data. This aspect is heavily impacted by the method of data preparation as well as the quality of the clock that realizes the UTC(k) timescale and its corresponding steering system. The author conducted extensive research focused on improving the data preparation process, exploring various methods to enhance the forecasting quality. These studies were presented in several works [1-12], and the findings indicate that further improvements in data handling can lead to even better results.

In the specific context of UTC(PL) forecasting, the method proposed by the author for generating UTC(k) scale forecasts using a GMDH neural network has been successfully applied. This

approach uses data from hydrogen masers and commercial caesium atomic clocks to implement the UTC(PL) timescale, aiming to achieve forecast accuracy comparable to the highest-performing timescales. Notably, the forecasting system has been successfully integrated into the Polish Timescale UTC(PL), confirming its potential for widespread application in national metrology institutes (NMIs). This approach allows for continuous improvement and refinement of the timescale forecasting process, making it more efficient and reliable in real-time operations.

The research conducted on the Polish Timescale (UTC(PL)) forecasting using the GMDH neural network (NN) has led to the development of a forecasting system for UTC(k) timescales, as described in the works of Sobolewski and others (2017). This system has been tested using the Polish Timescale, which is based on a VCH-1003M hydrogen maser, selected for its precision and stability among several available clocks at the GUM (Central Office of Measures). The goal of this work is to demonstrate that the forecasting system enables the UTC(PL) timescale to achieve a level of accuracy comparable to some of the best timescales, such as those implemented with hydrogen masers supplemented by caesium fountains.

The research focuses on two prepared time series, TS1 and TS2, for use with the GMDH NN, and compares the forecasted values against the BIPM-designated UTC values ( $x_b(t)$ ). The forecasting results for both time series showed a high level of accuracy, with discrepancies consistently within  $\pm 10$  ns, and for some periods, even below  $\pm 6$  ns. This indicates that the Polish Timescale is very close in quality to the best global timescales.

The article also elaborates on the two main factors that influence forecasting quality: the choice of neural network and the quality of the input data. A key component of the GMDH NN's success is its self-organizing nature, which allows it to adapt efficiently to new data and produce reliable forecasts in real-time, as seen in the ongoing UTC(PL) steering process. This method has been successfully applied in Poland, allowing for continuous updates to the national timescale, positioning it among the highest-performing timescales globally.

When developing the forecasting algorithm for UTC(k) using a GMDH neural network, the Polish Timescale was implemented with a commercial caesium atomic clock, achieving a stability of approximately  $10^{-14}$ . The author highlighted that leading NMI laboratories typically use hydrogen masers, often supplemented by caesium fountains, to implement UTC(k). Hydrogen masers offer superior short-term stability at the  $10^{-15}$  level compared to caesium clocks, which are better for long-term stability.

The research showed that the GMDH-based forecasting system is versatile, achieving high-quality predictions for both caesium and hydrogen maser-based timescales. This system has been integrated into the Polish Timescale UTC(PL), which has demonstrated exceptional accuracy, particularly since its application in 2016 at the GUM (Central Office of Measures) to control UTC(PL). The improvements were substantial, contributing to a timescale that is now ranked among the best globally.

The article aims to assess the effectiveness of the forecasting system applied to the Polish Timescale (UTC(PL)), based on a method proposed by the author using a GMDH neural network. This system is implemented on the VCH-1003M hydrogen maser, selected from several available clocks at the GUM (Central Office of Measures) in Poland. The research demonstrates that the developed forecasting system can achieve a level of precision for UTC(PL) comparable to the best global timescales, including those based on hydrogen masers and caesium fountains. The article emphasizes that the use of this system ensures high-quality timescale accuracy for UTC(PL), positioning it as one of the top-performing timescales worldwide.

The article describes the construction of the TS1 and TS2 time series, which are essential for the forecasting system based on the Group Method of Data Handling (GMDH) neural network. These time series were developed to improve the accuracy of UTC(PL) forecasting. The time series include phase time values ( $x_a(t)$ ) between the national time scale UTC(k) and the atomic clock, UTC - UTC(k) values ( $x_b(t)$ ), and additional values ( $x_{br}(t)$ ) representing the UTC<sub>r</sub> - UTC(k).

The research conducted from December 2020 to August 2023 focused on forecasting the Polish Timescale (UTC(PL)), which is based on the VCH-1003M hydrogen maser. The forecasting

used the prepared TS1 and TS2 time series and compared the predicted results with actual data published by BIPM.

The study demonstrated that both time series provided highly accurate forecasts, with discrepancies between the forecasted values ( $x_{bp}(t)$ ) and the BIPM data ( $x_b(t)$ ) being minimal, often within a few nanoseconds. This indicates that the Polish Timescale has remained consistently accurate, aligning closely with the best global timescales.

An essential part of the developed forecasting system based on the proposed procedure is the block dedicated to preparing input data for the GMDH neural network (NN) in the form of time series. The quality of the forecasts heavily depends on how this data is prepared. For the UTC(k) forecasting research, two time series, TS1 and TS2, were developed, containing data with a one-day interval.

The TS1 time series consists of three data groups: phase time values ( $x_a(t)$ ) between 1 pulse-per-second signals from the national UTC(k) timescale ( $UTC_k(t)$ ) and the atomic clock ( $clock_k(t)$ ),  $x_b(t)$  values representing the difference  $UTC - UTC(k)$ , and  $x_{br}(t)$  values representing the difference between  $UTC_r$  and  $UTC(k)$ . These relationships are defined as:

- $x_a(t) = UTC_k(t) - clock_k(t)$
- $x_b(t) = UTC(t) - UTC_k(t)$
- $x_{br}(t) = UTC_r(t) - UTC_k(t)$

The  $x_b(t)$  values are the data published by the BIPM in their "Circular T" bulletin. Since these data are published for Modified Julian Dates (MJD) ending in 4 and 9, they are interpolated using the Hermite polynomial (PCHIP function in MATLAB) to generate daily values. This interpolation helps extend the number of historical data points, ensuring the GMDH NN has enough training data, as insufficient data can hinder the training process.

The preparation method for TS2 is similar to TS1, but the data is split differently: the first group consists solely of  $x_b(t)$  values, while the second group consists only of  $x_{br}(t)$  values.

The forecasting process involves predicting  $x_b(t)$  values. Using TS1, the output from the GMDH NN for a forecast day ( $tp$ ) is denoted as  $x_{1p}(tp)$ . By comparing this with the measured  $x_a(tp)$  for  $UTC(k)$  on that day, the forecast difference ( $x_{bp}(tp)$ ) is calculated as:

- $x_{bp}(tp) = x_{1p}(tp) - x_a(tp)$

For TS2, the forecasting approach is similar, and the results can also be used to correct  $UTC(k)$  values.

## Conclusions.

The research demonstrates that the forecasting system based on the GMDH neural network provides highly accurate predictions for the Polish Timescale ( $UTC(PL)$ ), enabling it to achieve a level of precision comparable to the best global timescales. The system's use of hydrogen masers, along with well-prepared input data and the application of advanced forecasting techniques, ensures that the accuracy of  $UTC(PL)$  remains within a narrow range, often within  $\pm 10$  nanoseconds.

The developed system has proven to be effective in real-time forecasting and steering of  $UTC(PL)$ , which has significantly enhanced its performance. The approach has shown to be adaptable, with minimal discrepancies between forecasted and actual data, making it a reliable tool for maintaining precise timekeeping.

The research results indicate that the GMDH neural network is an optimal choice for forecasting  $UTC(k)$  timescales, especially for national metrology institutes (NMIs) that do not have access to high-cost equipment like caesium fountains. By using this system, such institutions can achieve high-quality time scale predictions at a significantly lower cost.

Furthermore, the application of this forecasting system has contributed to the Polish Timescale being classified among the top-performing timescales, demonstrating the practical benefits of artificial intelligence in metrology. The continued use of this system ensures the ongoing precision and reliability of  $UTC(PL)$ , positioning it as a leading example of modern timekeeping technology.

## REFERENCE

1. Miczulski, W., Sobolewski, L., 2011. *Application of the GMDH neural networks in prediction of corrections of the national time scale. Electronics: Constructions, Technologies, Applications* 6, 45–47.
2. Miczulski, W., Sobolewski, L., 2012. *Influence of the GMDH neural network data preparation method on UTC(PL) correction prediction results. Metrol. Meas. Syst.* 19 (1), 123–132.
3. Miczulski, W., Sobolewski, L., 2017. *Algorithm for predicting [UTC - UTC(k)] by means of neural networks. IEEE Trans. Instrum. Meas.* 66 (8), 2136–2142.
4. Miczulski, W., Sobolewski, L., Croonenbroeck, C., 2015. *A Neural model of a wind turbine. In: Proc. XXI IMEKO World Congress Measurement in Research and Industry*, pp. 1–4.
5. Sobolewski, L., 2016. *Application of the Neural Networks for Predicting the Corrections for the National Timescale UTC(PL)*, Zielona Gora, Poland: University of Zielona Gora. Press.
6. Sobolewski, L., 2017a. *Application of GMDH type neural network for predicting UTC(k) timescales realized on the basis of hydrogen masers. In: Proc. Joint IEEE (FCS EFTF)*, pp. 42–46.
7. Sobolewski, L., 2017b. „Predicting the Lithuanian Timescale UTC(LT) by means of GMDH neural network,”. *Bulletin of the Military University of Technology* 66 (4), 31–41.
8. Sobolewski, L., 2020. „Comparison of quality of UTC(PL) and UTC(NPL) scales prediction by means of GMDH neural network,”. *Measurement Systems In Theory and in Practice*, Zielona Gora, pp. 229–242.
9. Sobolewski, L., 2022. „Jako 's'c prognozowania skali czasu nadzorowanej przez pierwotny wzorzec częstotliwości z zastosowaniem sieci neuronowej typu GMDH na przykładzie skali czasu UTC(NPL),” . *Electr. Rev.* 11, 131–134 (in Polish).
10. Sobolewski, L., Miczulski, W., 2016. *Methods of time series preparation based on UTC and UTCr scales for predicting the [UTC-UTC(PL)]*. *J. Phys., Conf. Ser.* 723, 012040.
11. Sobolewski, L., Miczulski, W., 2021. *Methods of constructing time series for predicting local time scales by means of a GMDH-type neural network. Appl. Sci.* 11 (12).
12. Sobolewski, L., Miczulski, W., Czubla, A., 2021. *Experimental verification of the neural network predicting procedure applied for UTC(PL). IEEE Trans. Instrum. Meas.* 70 (Sep).

**Mateusz Tejer, Rafal Szczepanski, Tomasz Tarczewski** (*Institute of Engineering and Technology, Nicolaus Copernicus University, Torun, Poland*)

## **ROBUST TASK SCHEDULING IN ROBOTICS USING REINFORCEMENT LEARNING.**

***Abstract:** Effective task scheduling plays a crucial role in enhancing performance, productivity, and profitability across various real-world applications, including production lines, logistics, and transportation systems. While traditional scheduling methods often rely on heuristics or basic rule-based strategies, the rise of machine learning and artificial intelligence has spurred increased interest in leveraging these advanced technologies for task optimization. Among these, reinforcement learning stands out as a promising approach due to its ability to learn from experience and adjust to dynamic conditions. A frequently overlooked aspect, however, is the selection of optimal algorithm parameters and the diverse ways in which the environment can be modeled. This study evaluates the potential of reinforcement learning for task scheduling, offering a detailed analysis that aids in selecting the most effective environment models and Q-learning parameters. Additionally, the study introduces an automatic parameter selection method based on optimization algorithms. Despite these improvements, the system's adaptability to environmental changes remains limited. This observation led the authors to propose a novel Hybrid Q-learning approach, designed to ensure higher efficiency regardless of varying environmental factors.*

***Keywords:** task scheduling, performance, productivity, reinforcement learning, optimization, Q-learning, algorithm parameters, environmental adaptability*

### **Introduction:**

Effective task scheduling is crucial for enhancing performance, productivity, and profitability across various fields like manufacturing, logistics, and transportation systems. Traditionally, task scheduling has relied on heuristics and rule-based methods. However, with the growth of machine learning and artificial intelligence, there is increasing interest in using these technologies for optimization. Among these methods, reinforcement learning stands out because it can learn from experience and adapt to changing conditions. A key element often overlooked is the selection of optimal algorithm parameters and the modeling of the environment. This study evaluates how reinforcement learning can improve task scheduling, providing insights into efficient environment models and Q-learning parameters. Additionally, it proposes an automated selection process using optimization algorithms. Despite these advancements, the system's adaptability to environmental changes remains limited, motivating the authors to develop a Hybrid Q-learning approach that ensures superior efficiency under varying conditions.

In manufacturing, robotic arms are commonly used for various tasks, including pick-and-place operations, welding, painting, palletizing, and sorting. Traditionally, robotic arms are programmed manually due to a lack of advanced automatic algorithms that can deliver consistent, high-performance results across different processes. The system's performance, and consequently the quality of the product, largely depend on the skills of the operator. Additionally, these tasks often require collaboration between various machines, such as conveyor belts, mobile robots, and other robotic arms. Although creating a real-world testing environment for development can be costly, incorporating robotic arms into active production lines for experimentation is not a feasible option. As an alternative, simulation-based training offers a cost-effective way to generate the required data, though oversimplifying simulations can lead to discrepancies with real-world performance, resulting in unsafe behavior when the system is deployed in real robots. To reduce the complexity of robotic programming, tasks can be broken down into three sub-problems: task planning, task scheduling, and motion planning.

Task planning involves selecting the sequence of actions needed to accomplish a given task. Task scheduling determines the optimal order in which tasks are to be performed, while motion planning focuses on developing collision-free strategies for moving the robot to its target destination. Most task scheduling problems aim to improve production efficiency through better control or optimized velocity profiles. Task scheduling is commonly used in fields such as public transport, grid computing, and cloud systems. It can be categorized into static and dynamic scheduling; the former

involves pre-scheduling all tasks before execution, whereas dynamic scheduling adjusts decisions in real-time. Many task scheduling problems are formulated as variations of the Traveling Salesman Problem (TSP) or the Job Shop Scheduling Problem (JSSP), which are both NP-hard.

To address the task scheduling problem, machine learning techniques, particularly reinforcement learning (RL), offer a promising solution. Machine learning, a subset of artificial intelligence, enables systems to learn and improve over time without explicit programming, based on various types of data. Among the different types of machine learning, reinforcement learning stands out due to its ability to learn through interactions with the environment, unlike other methods that rely on predefined datasets. The agent in RL receives feedback in the form of rewards or penalties, depending on the actions it takes, and these interactions occur over multiple steps, called epochs. At the end of each epoch, the environment is reset to its initial state. RL has been applied successfully in diverse fields, such as energy system control, autonomous vehicles, traffic signal management, cybersecurity, and humanoid robots.

Reinforcement learning algorithms include several variations, such as Q-learning, Deep Q-learning, and Monte Carlo methods. Q-learning, in particular, is popular for its simplicity and ease of implementation. This algorithm uses a Q-table, which represents the expected future rewards for state-action pairs. However, the dimensionality of the Q-table increases exponentially with the complexity of the problem, making it less suitable for more complex tasks. Although implementing Q-learning is relatively straightforward, selecting optimal parameters for specific applications is not. The existing literature often fails to justify or even present the chosen parameters, which can affect the reproducibility of the results. Despite these challenges, RL continues to gain attention for solving complex scheduling problems, especially in environments that change dynamically.

In situations where the task's success rate is used as a performance metric, achieving 100% success is the maximum possible value. This threshold can serve as a stopping criterion for the learning process. However, when the performance metric involves efficiency, such as the number of products processed per hour, latency, or movement time, it becomes more challenging to determine whether the solution provided by an algorithm is optimal and when to halt the learning process.

This issue is particularly relevant to robotic sorting systems, where a robotic arm is tasked with selecting goods from three independent sources and sorting them into designated destinations. The use of reinforcement learning (RL) can enhance the efficiency of the robotic arm by addressing the task scheduling problem. The study focuses on analyzing the impacts of Q-learning parameters, environment models, and reward functions, with the goal of improving system performance. Predicting future states is also incorporated to further optimize the robotic sorting process.

In addition to the various combinations of algorithm parameters, choosing the right reward functions, resetting modes, input parameters, or neural network sizes can be just as important, if not more so. This paper presents a detailed analysis of these variables and proposes an automatic selection process using optimization algorithms. Despite achieving optimal parameters for a predefined environment, the solution may become inefficient when environmental conditions change. These findings prompted the development of a Hybrid Q-learning approach, which ensures consistent performance even when environmental factors fluctuate.

The proposed approach for task scheduling in robotic systems is based on an offline learning process using a highly accurate environmental model. By optimizing this model, it ensures that the efficiency of the robotic arm is maximized, even when environmental conditions change. The Hybrid Q-learning approach is particularly effective because it can maintain high performance without requiring constant adjustments during the operational phase, as opposed to traditional online learning methods.

Experiments have shown that the Hybrid Q-learning approach significantly outperforms both static scheduling and online learning algorithms. The offline learning process allows the model to adapt to changes in the environment more efficiently, ensuring that the scheduling remains optimal over time. In comparison, online learning methods require longer adaptation periods, during which the performance is often lower.



The paper also demonstrates the resilience of this approach, showing that even after significant environmental changes, the system can quickly adapt and continue to perform effectively. This is achieved through the use of a pre-tuned Q-table that is capable of handling environmental variability without losing efficiency.

This method not only provides a robust solution to task scheduling in robotic applications but also offers a path forward for future enhancements, such as incorporating multi-agent systems and improving motion planning and collision avoidance algorithms.

Two experiments were conducted to demonstrate the superiority of the proposed Hybrid Q-learning approach. The first experiment involved the following steps: (i) optimization of the scheduling using Q-learning based on the nominal parameters of the sorting stand, (ii) after 2 hours, the environmental parameters were changed to those corresponding to Environment #2, (iii) a 12-hour relearning period was initiated, and (iv) the final performance was measured using the learned Q-table.

In the second experiment, the environment parameters were changed after 2 hours, with a one-hour delay in the changes. The proposed approach was compared to static scheduling (which does not adjust the task scheduling) and online learning, where either the previous Q-table was used as a starting point (Online #1), or no prior knowledge was available (Online #2). The comparison involved calculating the number of sorted products over a sliding window of 250 products, which represents approximately 30 minutes of operation.

In the first experiment, keeping the scheduling fixed resulted in a decrease in performance, while the Hybrid Q-learning approach demonstrated superior performance. The Online learning methods, especially without prior knowledge, were found to be less efficient and resulted in suboptimal scheduling after the learning period. These results confirm that the Hybrid Q-learning approach significantly improves the efficiency of task scheduling in dynamic environments, especially when changes in environment parameters occur during the task execution phase.

To assess the efficiency of the scheduling algorithms, the number of sorted products was calculated as the average over a sliding window of 250 previously picked products, which corresponds to approximately 30 minutes of operation. In the first experiment, where no change was made to the scheduling, the results showed that keeping the scheduling static led to reduced performance. On the other hand, the Hybrid Q-learning approach demonstrated consistently higher efficiency. In the second experiment, where environment parameters were altered after two hours, the proposed Hybrid Q-learning approach still outperformed static scheduling and online learning methods, with or without prior knowledge of the system's Q-table.

The comparisons reveal that the Hybrid Q-learning approach was particularly advantageous when the environment changed, as it was able to adapt faster and achieve higher performance than other methods. The system's performance was tracked across various iterations, showing the difference in the number of products sorted per hour. Notably, the Hybrid Q-learning method delivered superior results by allowing the system to adapt during robotic arm movements, thus optimizing task scheduling in real-time.

In contrast, online learning methods—especially those starting from scratch (Online #2)—exhibited suboptimal performance due to their slow adaptation to environmental changes. Moreover, these methods required additional time for learning, reducing their effectiveness in rapidly evolving scenarios. The static scheduling approach, while more efficient than online methods in some cases, could not adapt to changes in the environment and therefore was outperformed by the more dynamic Hybrid Q-learning solution.

These findings highlight the superior efficiency and flexibility of the Hybrid Q-learning approach in dynamic environments, making it a promising solution for task scheduling problems where environmental conditions are subject to change.

The paper highlights the significance of selecting an appropriate environment model with predictive capabilities and tuning the parameters of the Epsilon-Greedy Q-learning algorithm for effective task scheduling. Three prediction strategies, four reward functions, three resetting modes, and various robot arm history positions were analyzed to determine the best set of parameters,

resulting in nearly 31 million possible combinations. The study demonstrates that when the environment's behavior changes significantly—especially when predictive modes are involved—there can be a drastic shift in the optimal Q-learning parameters. This emphasizes the need to regularly update the Q-learning parameters to maintain efficiency under varying conditions.

Based on these findings, the authors propose a Hybrid Q-learning approach, which ensures high efficiency from deployment to real-world environments by adapting the Q-table using simulation learning throughout robotic arm operations. Experiments show that online relearning methods require much more time to adapt, resulting in lower performance during the learning period. In contrast, the proposed Hybrid Q-learning approach adjusts immediately, maintaining superior performance even under changing conditions.

Looking ahead, the authors plan to expand this approach to handle multiple robotic arms and address challenges like motion planning, collision avoidance, and multi-agent systems.

The authors plan to extend their work by focusing on task scheduling in environments with multiple robotic arms. This future research will explore additional challenges, such as motion planning, collision avoidance, and the application of multi-agent algorithms. These improvements aim to further enhance the efficiency and adaptability of task scheduling systems in complex, dynamic environments.

**Mathias Held, Jannis Bulling, Yevgeniya Lugovtsova, Jens Prager** (*Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany*)

## **ELASTIC CONSTANTS FROM ULTRASONIC DISPERSION IMAGES VIA NEURAL NETWORKS.**

**Abstract:** *This paper introduces a method for determining the isotropic elastic constants of plate-like structures by leveraging the dispersive characteristics of ultrasonic guided waves combined with neural networks. This is achieved through the analysis of dispersion images. Two distinct neural network architectures are compared: one utilizing convolutions and transfer learning based on EfficientNetB7, and another inspired by Vision Transformer approaches. To support this, both simulated and experimental dispersion images are created, with the former being used for designing, training, and validating the networks, while the latter serves for testing. During the training phase, various data augmentation techniques are applied to the simulated data to replicate the artifacts present in the measured data, enabling the networks to generalize from simulations to real-world measurements. The performance of the trained neural networks is evaluated on dispersion images from seven different material samples, testing several variations of the measured images to ensure prediction consistency. The study highlights that neural networks can successfully predict isotropic elastic constants from experimental dispersion images, relying solely on simulated images for training and validation, without requiring an initial estimate or manual feature extraction, and regardless of the measurement setup. Additionally, the paper discusses the effectiveness of the different architectures for extracting information from dispersion images and introduces a novel image-to-regression visualization technique.*

**Keywords:** *ultrasonic, guided waves, neural networks, dispersion images, isotropic elastic constants, transfer learning, data augmentation, prediction stability*

### **Introduction:**

The article discusses a method that utilizes the dispersive behavior of ultrasonic guided waves (UGWs) along with neural networks to determine the isotropic elastic constants of plate-like structures through dispersion images. The authors compare two neural network architectures: one based on convolutional layers with transfer learning using EfficientNetB7, and another using a Vision Transformer-like approach. To implement this method, both simulated and measured dispersion images are generated. Simulated images are used to design, train, and validate the neural networks, while the measured images are used for testing.

During the neural network training, specific data augmentation layers are applied to introduce artifacts typical in measurement data into the simulated data. These layers allow the neural networks to extrapolate from simulated data to real-world measurements. The neural networks are trained on images from seven known materials, and the study tests various variations of the measured images to ensure the stability of predictions. The results show that the neural networks can predict the isotropic elastic constants from measured dispersion images using only simulated data for training and validation, without the need for an initial guess or manual feature extraction, regardless of the measurement setup. The paper also discusses the suitability of the different architectures for extracting information from dispersion images and presents an image-to-regression visualization technique.

This approach promises automation of the material characterization process by eliminating the need for initial guesses or manual data extraction, and it is applicable across different measurement setups.

The paper utilizes a semi-analytical numerical simulation method, known as the Scaled Boundary Finite Element Method (SBFEM), to simulate dispersion images for structures with defined elastic constants. These simulated dispersion images are generated by solving the eigenvalue problem of the SBFEM equation for specific wavenumbers, resulting in corresponding frequency values. The isotropic elastic constants are evenly distributed within the dataset to prevent neural networks from overfocusing on any particular range. A total of 20,000 dispersion images are generated, covering the specified range of values, as shown in Table 1. Additionally, a dataset containing 2,000 dispersion images is randomly distributed for testing purposes.

The trajectory of the modes of ultrasonic guided waves (UGWs) within a dispersion image in the frequency-wavenumber domain is directly proportional to the thickness of the plate. To simplify the simulation process and facilitate comparison across different plate thicknesses, the axes are normalized by thickness. This method enhances the robustness of the material characterization by making the comparison more straightforward. In these simulations, instead of directly using Young's modulus and density, two normalized quantities are applied:

$$E_{norm}=E/\rho \quad \text{and} \quad \rho_{norm}=1$$

where  $\rho_{norm}$  and the thickness are set to one. This leaves  $E_{norm}$  and Poisson's ratio as the only parameters to define the shape of the curves, which the models can learn and predict from the dataset/

The measurement data was recorded using the Verasonics Vantage 64 system, which has a sampling frequency of 62.5 MHz. A phased array system, the Olympus 2.25-L-64-48X12-A2-P-2.5-HY with 64 elements, is employed, where the first element excites the ultrasonic guided wave (UGW) and the remaining 63 elements record it. The resolution along the frequency-wavenumber axis is limited by the small number of elements and their minimal spacing, especially considering the axis scaling by thickness in the dispersion images. The thickness of the specimen, a  $300 \times 300$  mm plate, is measured at each corner using a micrometer screw gauge, and the average value is used to calculate the material's density.

The data is processed according to the procedures described in previous studies. Initially, a Tukey window function (with  $\alpha = 0.3$ ) is applied to the time and space data. The data is zero-padded to the next larger power of two, ensuring it has at least 213 entries. The 2D Fast Fourier Transform (FFT) is then applied, shifting the zero frequency to the center. Following this, the absolute values are taken, and a quarter section of the dispersion image with positive group velocities is selected. The image undergoes gamma correction and normalization before a 21-time blur filter smooths the gradients within the pixels. An adaptive threshold function from OpenCV2 is applied, and the image is eroded and dilated multiple times to remove noise and adjust the image for further processing, ultimately resizing it into a  $600 \times 600$  binary array suitable for training the neural networks.

The neural networks are trained using both simulated and real-world data. To enhance the training process, data augmentation layers are incorporated into the models. These layers randomly apply augmentations to the simulated data, ensuring unique inputs during each training epoch. This technique introduces artifacts typical of measurement data into the simulated dataset, allowing the neural networks to learn how to extrapolate from simulated to measured data. The augmented images simulate various measurement-related anomalies such as noise and spectral leakage, improving the model's robustness to real-world variations.

It is crucial that these augmentations are integrated directly into the training process using TensorFlow's syntax, enabling faster processing with GPU computing. Through these augmentations, the model becomes more capable of handling imperfections in the measurement data, such as variations in the mode shapes or artifacts caused by measurement noise. These techniques are shown to improve prediction accuracy by making the neural networks more resilient to real-world challenges during the testing phase.

The article explores the use of two advanced neural network architectures for predicting the isotropic elastic constants from dispersion images of ultrasonic guided waves (UGWs) in plate-like structures. The two architectures compared are EfficientNetB7, a convolutional neural network (CNN) with transfer learning, and a Vision Transformer model, both adapted for a regression task. The networks are trained using simulated dispersion images and validated with real-world data.

A series of data augmentation techniques are employed to simulate measurement artifacts in the training data, such as noise, dilation, and missing information, ensuring the neural networks can generalize well when faced with imperfections in the measured data. The models are trained with TensorFlow, utilizing a combination of Keras-Tuner and Hyperband for hyperparameter optimization. The performance of these architectures is evaluated using seven different material samples, with results showing that the Vision Transformer model tends to be more robust to variations

in the dispersion images, while the EfficientNetB7 model is more sensitive to minor changes in the data.

The study demonstrates the potential of neural networks for accurately predicting elastic constants without the need for manual feature extraction or initial guesses, which is a significant advantage over traditional methods that rely on explicit mathematical models. The article also discusses how neural networks, especially CNNs and Vision Transformers, are capable of handling complex, multi-modal information inherent in dispersion images, offering an automated and efficient approach for material characterization.

The study focuses on using neural networks to predict isotropic elastic constants from dispersion images, derived from ultrasonic guided wave propagation in plate-like structures. The models analyzed include EfficientNetB7 and a Vision Transformer model. Both architectures were adapted from classification tasks to solve this multi-output regression problem.

The training process utilized a mix of simulated and measured dispersion images, with significant data augmentation techniques to mimic real-world imperfections, such as noise, dilation, and spectral leakage. These augmentations helped the models generalize better, improving their ability to predict material properties despite artifacts in the data.

The EfficientNetB7 model, based on convolutional layers, converged faster due to pre-trained weights but was more sensitive to small changes in the input images. On the other hand, the Vision Transformer model showed greater robustness to variations and outperformed the EfficientNet model in prediction stability, particularly in the presence of measurement noise and artifacts.

Additionally, attention mapping techniques, such as Grad-CAM, were adapted for this regression task to visualize how the models made predictions. This method revealed that the Vision Transformer model was more capable of utilizing positional information and handling input variations more effectively than the EfficientNetB7 model.

The models' performance was tested on a range of materials, showing that both architectures could predict elastic constants with reasonable accuracy. However, the Vision Transformer model exhibited superior consistency and robustness, suggesting that it is better suited for this type of problem where positional information within the dispersion images is crucial. This approach, utilizing neural networks for dispersion image analysis, presents a promising method for automated material characterization in various industrial applications, especially in the presence of measurement noise and variations.

The results of the study indicate that both EfficientNetB7 and Vision Transformer models show high accuracy in predicting isotropic elastic constants, with Mean Absolute Percentage Error (MAPE) for the simulated dataset at around 2.94% for EfficientNetB7 and 2.13% for Vision Transformer, respectively.  $R^2$  values were also high, demonstrating strong model performance. In particular, the Vision Transformer model was found to be more robust when predicting the elastic constants from the measured dispersion images, as it demonstrated better consistency in the presence of noise and variations.

During testing on measured datasets, the models showed some variations in their predictions depending on the image alterations. The addition of noise and small artifacts in the dispersion images had a more significant impact on the EfficientNetB7 model, which was more sensitive to such changes. However, the Vision Transformer model was less affected by these variations, demonstrating its superior generalization ability in handling noise and measurement artifacts. This robustness is thought to be due to the Vision Transformer's capacity to analyze global relationships in the image, as opposed to the convolutional layers of EfficientNetB7 that focus on local features.

The study highlights the importance of diverse test data for evaluating neural networks, as models trained on simulated data may perform differently on real-world measurements with additional noise and artifacts. Both models, despite their differences in architecture, provided similar accuracy when predicting elastic constants for the material samples, with Vision Transformer slightly outperforming the EfficientNetB7 model in prediction consistency.

These findings suggest that neural network models, particularly the Vision Transformer, can be effective tools for material characterization, capable of predicting isotropic elastic constants with

minimal pre-existing knowledge or manual feature extraction. The approach is promising for automated applications, and the methods developed could be extended to the analysis of more complex, anisotropic materials in future studies.

The study introduces an innovative approach for determining isotropic elastic constants using neural networks to process dispersion images obtained from ultrasonic guided waves (UGWs). This method leverages the dispersive behavior of UGWs and applies two different neural network architectures: EfficientNetB7, a convolutional network with transfer learning, and a Vision Transformer-like model. These architectures are tested on both simulated and measured dispersion images, with the simulated data being used for training and validation, and the measured data for testing.

The training process incorporates various data augmentation techniques to introduce real-world artifacts, such as noise and spectral leakage, into the simulated data. This allows the neural networks to generalize better, bridging the gap between the simulated and measured data. The models are evaluated on seven different material samples to ensure prediction stability across a wide range of real-world scenarios.

The results show that both models can accurately predict the isotropic elastic constants from measured dispersion images, even when trained exclusively on simulated data. This method eliminates the need for initial parameter guesses or manual feature extraction, offering an automated solution that works independently of the measurement setup. The study also discusses the strengths of each architecture in handling complex image data, with the Vision Transformer showing greater robustness in predicting elastic constants across various conditions.

**Pawel Majewski, Mariusz Mrzyglód, Piotr Lampa, Robert Burduk, Jacek Reiner** (Wrocław University of Science and Technology, Wrocław, Poland)

## MONITORING INSECT LARVAE GROWTH WITH REGRESSION CNN AND KNOWLEDGE TRANSFER.

**Abstract:** *Insect larvae breeding, particularly *Tenebrio molitor* and *Hermetia illucens*, has become increasingly popular. Monitoring their size distribution and growth over time is essential, yet challenging due to the large number of overlapping larvae in images. To address this, we propose an efficient method for determining larval size distribution using a regression convolutional neural network (RegCNN) and knowledge transfer. We focused on larval width as the main parameter, as it is easier to measure in dense scenes, with length (L) and volume (V) estimated indirectly through regression models. The RegCNN was trained using knowledge transfer to minimize the time-consuming process of labeling numerous images. The width quartiles were determined using an enhanced multistage phenotyping process based on computer vision and a segmentation model. This approach required labeling only a few images for calibration. We evaluated several RegCNN architectures, including pre-trained models (ResNet, EfficientNet) and a custom model. The best model, ResNet18, achieved an RMSE of 0.131 mm for larval width and 1.12 mm for length, with an  $R^2$  of 0.870 and an average inference time of 0.30 seconds. The custom architecture, TenebrioRegCNN\_v3, was slightly less accurate but five times faster in inference time. The results demonstrate the method's potential for real-world breeding applications.*

**Keywords:** *insect larvae, size distribution, regression, convolutional neural network, knowledge transfer, segmentation, phenotyping, real-time monitoring*

### Introduction:

In recent years, there has been a surge in the popularity of breeding insect larvae, particularly *Tenebrio molitor* and *Hermetia illucens*. A crucial aspect of managing insect larvae breeding is monitoring their growth and size over time. However, determining the distribution of larvae sizes in real-time presents a significant challenge, particularly due to the dense scenes in images where larvae often overlap. To tackle this, we have proposed an efficient method for estimating the size distribution of larvae. This approach leverages a regression convolutional neural network (RegCNN) combined with knowledge transfer. The primary parameter measured was the larval width, as it is easier to detect in dense scenes. Larval length (L) and volume (V) were estimated indirectly using regression models for these parameters based on width.

The training of the RegCNN model was enhanced using knowledge transfer, which minimized the need for extensive labeling of images with larvae at various stages of growth. For the training process, we used width quartiles—lower quartile, median, and upper quartile—derived from a multistage phenotyping approach. This method relied on classical computer vision techniques along with a larvae segmentation model. This approach only required labeling a small subset of images for calibration purposes. The study also explored different RegCNN architectures, both pre-trained on ImageNet (such as ResNet and EfficientNet) and custom models with fewer parameters. The proposed method was validated by evaluating larvae distributions characterized by width quartiles ranging from 1.7 mm to 3.1 mm, which correspond to average larval lengths between 16 mm and 28 mm.

This method showed promising results, especially with the ResNet18 model, which achieved a root mean square error (RMSE) of 0.131 mm for larval width estimation and an average RMSE of 1.12 mm for larval length estimation. The  $R^2$  value for the model was 0.870, with an average inference time of 0.30 seconds per box. The custom architecture, TenebrioRegCNN\_v3, had slightly lower accuracy but was about five times faster in inference time per image. The results validate the potential of this method for real-world applications in insect larvae breeding.

The acquisition of images from insect breeding boxes was carried out under industrial breeding conditions, using a machine vision system placed on an automated robot servicing the breeding process. The images were captured with a color camera (GOX-12401C, JAI, Denmark) with a resolution of 4096 × 3000 pixels. The camera was positioned at a distance that allowed for imaging the entire surface of the box, with a resolution of 0.143 mm/pixel. The area was illuminated using

cool white LED strips to minimize the impact on the insects during image capture. The raw images were preprocessed to compensate for shading caused by insufficient lighting and to correct lens distortion using a calibration method based on a gray pattern and a chessboard pattern.

For the phenotyping of larvae, the process involved segmenting the larvae from the images using an instance segmentation model. The segmentation was followed by determining the size parameters for each larva using classical computer vision techniques. Specifically, the width of each larva was determined from the segmentation mask, and this measurement was used to calculate other parameters such as length and volume indirectly using linear regression models. These models were derived from larvae in the dataset D1, with 266 points used for the regression analysis.

To address challenges in segmenting individual larvae from densely packed scenes, synthetic images were generated by placing individual larvae images onto a background. This approach was essential for training the instance segmentation model, reducing the need for extensive manual annotation.

This method of multistage phenotyping was crucial for training the regression convolutional neural network (RegCNN), which allowed for the prediction of larval size quartiles directly from input images. This solution significantly reduced the need for manual annotation, requiring only a few images for calibration purposes. The use of synthetic images and knowledge transfer between multistage phenotyping and RegCNN training allowed for faster and more accurate predictions in real-world applications.

The improved multistage phenotyping method used in this study for larval size determination relies on classical computer vision techniques. This method serves two purposes: it provides pseudo target values for training the regression convolutional neural network (RegCNN) and helps define the linear regression models for larval length and volume based on width measurements. The term "pseudo target values" refers to the size estimates generated through this method, while "true target values" are those derived from manually annotated larvae images.

Phenotyping can be applied either to a population of larvae within a single breeding box or to individual larvae. For population phenotyping, the process involves determining the size parameters such as the lower quartile (Q1), median (Q2), and upper quartile (Q3) for width, length, and volume. In the case of individual larvae, the process begins with segmenting the larvae from the image using an instance segmentation model. The next step is to determine the size parameters for each segmented larva using classical computer vision methods.

The segmentation process provides a binary mask, which precisely defines the larva's location in the image. For phenotyping, two sets of pixels are defined: those in the binary mask and those along the contour of the mask. Phenotyping involves several steps, including the creation of a smoothed skeleton of the larva and the calculation of the width based on sections perpendicular to the skeleton. The method, which avoids issues found in previous skeletonization techniques, ensures that the calculated larval width is accurate and robust, even for small or highly curved larvae.

To calculate the larval size parameters, a linear regression model was developed, establishing relationships between the width of the larva and its length and volume. This approach allows the indirect calculation of length and volume based on the measured width. The regression models were trained using larvae extracted from the D1 dataset, providing a reliable basis for determining the size parameters across different larvae populations.

The segmentation process itself involved an iterative improvement of the instance segmentation model. Starting with a small dataset of synthetic larvae images, the model was progressively refined by adding real data and increasing the diversity of the training set. This multistage development process led to a significant improvement in the model's ability to segment larvae accurately across a range of sizes and conditions.

Finally, the use of this improved multistage phenotyping method in conjunction with the RegCNN for predicting larval size quartiles enables real-time, scalable phenotyping of larvae populations in industrial breeding environments. This hybrid approach, combining classical image



processing techniques with deep learning, provides a powerful solution for efficient monitoring of larval growth and development.

The width of the larvae was chosen as the main and directly measurable parameter due to its ease of registration in dense scenes. However, from the breeder's perspective, the length is easier to assess. Conversely, when assessing the volume or mass gain of the larvae during their growth, volume becomes the more appropriate parameter. To enable the indirect calculation of length and volume based on the measured width of the larvae, linear regression models for length (as a function of width) and volume (as a function of width) were established.

These regression models were derived using larvae images from the D1 dataset, resulting in 266 data points for the regression analysis. The process included determining larval width as the median of the lengths of sections perpendicular to the skeleton, which was defined by the binary mask from the segmentation process. Volume was approximated by the sum of cylinder volumes, where each cylinder represented a section of the larva with its length and width at the corresponding point of the skeleton. A correction factor, determined experimentally, was applied to refine the calculations of volume. This approach enabled the estimation of larval length and volume indirectly from the width, leveraging the relationships established in the linear regression models.

This methodology provides a robust approach for calculating key size parameters (length and volume) based on a simpler and more reliably measured parameter, larval width, and supports efficient monitoring of larval growth in breeding environments.

The process of extracting individual larvae images and generating synthetic images plays a critical role in the multistage phenotyping of larvae. The segmentation of larvae is carried out using the Mask R-CNN instance segmentation model, which requires a set of labeled images. Labeling real images is highly time-consuming, especially due to the dense scenes, where larvae overlap significantly. To address this challenge, synthetic images with automatically generated labels are used. The generation of synthetic images involves selecting larvae from a prepared pool of individual larvae images and randomly placing them onto a background. This process is parameterized by the degree of overlap between larvae instances and the number of larvae to be included in the image.

The pool of larvae instances used for generating synthetic images consists of larvae extracted from the D1 dataset, which includes 266 individual larvae. These synthetic images are then used to train the initial larvae segmentation model. The training process allows the model to improve its ability to accurately segment larvae from new images, even in dense scenes. This approach, described in detail in [1], model's performance. The synthetic image generation process provides a foundational dataset for developing robust instance segmentation models.

In the study, synthetic images were used to train the initial model, which was subsequently tested and improved through several stages. The use of synthetic images and their automatic labeling significantly reduced the time and effort required for manual annotation while enhancing the model's ability to handle dense scenes efficiently. This approach allowed for the development of a larvae segmentation model capable of being applied in real-world industrial breeding conditions, where larvae are often densely packed and difficult to segment manually.

The larvae segmentation model was developed using a three-step approach, which sequentially improved the model's performance. The first step involved training the model on 200 synthetic images created from 266 larvae instances. These images were generated using larvae extracted from the D1 dataset. In the second step, the trained model was used to infer larvae from the D3.TRAIN dataset, yielding larval mask proposals for 489 images. This stage increased the object pool size to 65,000 larvae instances, and 1,000 synthetic images were generated for further model training. The third step involved training the model on both real images and the synthetic data from step two, treating the model's predictions as pseudo labels for the real images.

This approach aimed to minimize user input during training and maximize efficiency by using a combination of synthetic and real images. The segmentation model was trained with progressively larger datasets to ensure its robustness in various breeding conditions. The resulting model showed a significant improvement in larvae segmentation accuracy as the size of larvae increased, with the model performing better for larger larvae (27–35 mm). The segmentation accuracy was assessed

using AP50 and F1 scores, where the model showed sufficient accuracy for effective multistage phenotyping, particularly in detecting and measuring larvae within breeding boxes.

To address the issue of detecting small objects with deep learning models, a correction factor was introduced to adjust the calculated larval width quartiles. This factor compensates for variations in detection accuracy depending on the larval size. It was expected that the correction factor would decrease as the larval width increased. A diagram illustrating this correction factor as a function of larval width is presented in the text.

The correction factor was calculated based on the D2 dataset, which included manually labeled larvae instances. Histograms of larval widths were created from both the ground truth data and the predictions generated by the larvae segmentation model. The correction factor for each width interval was determined by dividing the histogram value of the ground truth by the predicted value. Special handling was applied to cases where no predictions were made for a specific width, setting the correction factor to 1 in these cases. The factor was adjusted at the boundaries of the width range to maintain consistency.

This correction was necessary to improve the accuracy of the quartile values, especially for smaller larvae that were often underdetected by the segmentation model. For widths below 2.0 mm, the correction factor was higher due to the model's difficulty in detecting these smaller larvae. The correction helped to minimize the overestimation of quartile values for these smaller larvae, improving the overall accuracy of the model when estimating size parameters. The calibration curve for the correction factor, as well as the impact on the final quartile calculations, is also discussed in the text.

In this study, the main challenge addressed was the knowledge transfer between improved multistage phenotyping based on classical computer vision methods and a regression convolutional neural network (RegCNN). Knowledge transfer involved training the RegCNN on values derived from the multistage phenotyping of larvae samples from the D3.TRAIN dataset. For each sample, the lower quartile (Q1), median (Q2), and upper quartile (Q3) of larval width were determined. Correction weights were incorporated during the calculation of quartiles, taking into account observations from earlier sections.

The RegCNN was trained using these quartile values, allowing for the direct estimation of larval width quartiles without analyzing individual larvae separately. The input to the RegCNN consisted of 800×800 RGB images, and the model outputted the values for the three quartiles of larval width. The training process minimized the mean squared error (MSE) loss, with the D3.TRAIN dataset used for training. Pre-trained convolutional neural networks (CNNs) such as ResNet and EfficientNet, along with custom architectures, were evaluated.

This approach successfully reduced the manual annotation effort, requiring only a few manually labeled images for calibration. The final model achieved high accuracy with minimal inference time, demonstrating its potential for real-time applications in large-scale insect larvae breeding systems. The use of synthetic images for training the segmentation model further contributed to the efficiency of the approach.

The regression convolutional neural network (RegCNN) developed in this study enables direct estimation of larval width quartiles (Q1, median, Q3) from 800×800 RGB images without requiring individual larvae analysis. The model outputs the three quartiles for larval width based on the input image. During training, the mean squared error (MSE) was minimized, using images from the D3.TRAIN dataset, which were resized from 1024×1024 to 800×800 pixels before input.

Pre-trained deep convolutional neural network architectures, such as ResNet18, ResNet50, ResNet101, EfficientNet-b0, EfficientNet-b4, and MobileNetv2, were evaluated. Custom CNN architectures with reduced complexity were also proposed. Fine-tuning was applied to all model weights, including the convolutional and fully connected (FC) layers. For pre-trained models, the FC layers were customized depending on the number of neurons in the input layer, with ReLU activation functions applied between layers.

The custom RegCNN architecture consisted of convolutional blocks followed by an FC block, where the number of neurons in the hidden layers varied depending on the size of the input layer.

Training parameters included a learning rate of 0.001, 200 epochs, and an Adam optimizer. The 'batch size' was adjusted based on the GPU capabilities. Pre-processing involved standardizing pixel intensity values in the R, G, and B channels according to ImageNet standards.

This setup aimed to efficiently estimate larval width quartiles while minimizing inference time. The results showed that the RegCNN could be effectively trained using a knowledge transfer mechanism from multistage phenotyping, requiring only a few manually labeled images for calibration. This method demonstrated potential for real-time applications in large-scale insect larvae breeding systems.

The larvae segmentation model's performance was evaluated in terms of the average precision at 50% intersection over union (AP50) and F1-score across different larval size sets. The results showed consistent improvements across successive stages of model training. Initially, when trained only on synthetic images, the model achieved an AP50 of 75.0%, which improved to 79.2% after further refinement with both synthetic and real images. The greatest improvement was observed in larvae with smaller sizes (18–23 mm), where the AP50 increased from 61.7% to 72.1%, while for larger larvae (27–35 mm), the model's accuracy remained high even in the first stage.

In the context of multistage phenotyping, the model's performance was highly sensitive to the larval size, with better results for larger larvae. This observation highlights the necessity of adjusting the calculation of quartiles for larval width in dense scenes, where smaller larvae may be underrepresented or misclassified. These improvements were sufficient to meet the needs of large-scale breeding systems, where quick and accurate larvae size monitoring is essential. The overall system demonstrated robust performance with an acceptable trade-off between accuracy and inference time, especially when processing large numbers of larvae in real-time.

The final validation of the model's performance was carried out using a variety of metrics, including the average precision at 50% intersection over union (AP50) and the F1-score. These metrics were calculated for different larval size sets. The results showed significant improvement across training stages. In the first stage, where only synthetic images were used, the model achieved an AP50 of 75.0% on average. After including real images in the second stage, the AP50 improved to 75.8%, with the most substantial increase observed for smaller larvae (18–23 mm), where the AP50 increased from 61.7% to 72.1%. The largest larvae (27–35 mm) showed high accuracy from the beginning, with an AP50 of 86.2%.

The final stage of training, which combined both real and synthetic images, led to an average AP50 of 79.2% and an F1-score of 0.824. For the larval size subset of 27–35 mm, the AP50 reached 86.0%, demonstrating the model's strength in detecting larger larvae. These results confirmed the effectiveness of the segmentation model for multistage phenotyping and emphasized the need for adjustments when calculating quartiles for smaller larvae due to segmentation challenges. The model was able to detect sufficient larvae instances, enabling accurate size parameter estimation for the larvae population, even in dense scenes. The achieved metrics were sufficient to support real-time applications in large-scale insect larvae breeding systems.

The analysis of inference time for the entire pipeline showed significant improvements with the proposed solution. The CNN-based regressor had the shortest inference time at 0.30 seconds per box, making it ideal for large-scale breeding applications where quick processing is crucial. In comparison, the multistage phenotyping method, when analyzing all larvae (approximately 1,300 per box), had a longer inference time of about 11 seconds per box. The second version of multistage phenotyping, which limited the number of analyzed larvae to 50 and reduced the area of the analyzed boxes to 25%, improved inference time to approximately 2.1 seconds per box.

In contrast, the reference method, based on the previous multistage phenotyping technique, had significantly higher processing times, with inference times reaching up to 217 seconds per box. This indicates that the proposed CNN-based regressor offers a considerable advantage in terms of efficiency, especially in scenarios involving large numbers of boxes (>10,000).

The inference time breakdown for the approaches considered revealed that for the CNN regressor with ROI relevance assessment, 33% of the total time was spent on extracting individual larvae using YOLOv5m, while 63% and 85% of the time for the two versions of multistage

phenotyping were spent on larvae segmentation using Mask R-CNN. These results highlight the need for further improvements in the segmentation step, particularly for Mask R-CNN, which is currently the bottleneck in the multistage phenotyping approach. Future work may involve replacing Mask R-CNN with a faster model, such as YOLOv8, to reduce processing times and improve overall efficiency.

The study demonstrated the effectiveness of a regression convolutional neural network (RegCNN) for determining the size parameters (width, length, and volume) of insect larvae, particularly *Tenebrio Molitor* and *Hermetia Illucens*. By applying knowledge transfer between improved multistage phenotyping methods and the RegCNN, the need for extensive manual annotation was significantly reduced. The multistage phenotyping method provided pseudo target values used for training the RegCNN, while linear regression models for length and volume were derived from larval width.

The performance of the CNN regressor was validated using datasets with both pseudo target values and manually annotated true target values. Results showed that the model achieved a high coefficient of determination ( $R^2 = 0.870$ ), with a root mean squared error (RMSE) of 0.131 mm for larval width. The proposed method also offered a fast inference time of 0.30 seconds per breeding box, making it suitable for real-time monitoring in large-scale breeding operations.

The study's findings indicate that this CNN-based solution strikes a balance between accuracy and processing speed, making it a promising tool for the large-scale, automated monitoring of insect larvae growth. Further work may focus on optimizing segmentation models and improving processing times for even faster and more efficient applications in commercial breeding systems.

The study outlines several important directions for future work. These include the development of models specifically tailored for inference in situations with very low larval densities, which would address current limitations in performance when larvae are sparse. Additionally, further reduction in the computational time for multistage phenotyping is a priority. This would allow this approach to be incorporated into hybrid phenotyping methods, improving efficiency.

Another key area for future development is the creation of methods for larval segmentation that can achieve similar accuracy across a broader range of larval sizes. Additionally, exploring amodal segmentation models, which can estimate missing or occluded larvae segments, is another promising direction. There is also potential in developing reference models for larval growth under controlled feeding conditions, which could support more accurate monitoring and growth prediction.

Finally, the development of methods for detecting anomalies in larvae growth, based on reference models, is another critical area for future research. These methods would be crucial in identifying irregularities in growth patterns. Furthermore, maintaining and adapting the proposed methods in the face of changes in data characteristics (domain shift) remains an important challenge for ongoing research.

## REFERENCES

1. Majewski, P., Zapotoczny, P., Lampa, P., Burduk, R., Reiner, J., 2022. Multipurpose monitoring system for edible insect breeding based on machine learning. *Sci. Rep.* 12 (1), 1-15.

**Tamara Sobot, Vladimir Stankovic, Lina Stankovic** (University of Strathclyde, United Kingdom)

## **HUMAN-IN-THE-LOOP ACTIVE LEARNING FOR TIME-SERIES ELECTRICAL DATA.**

***Abstract:** Advanced machine learning models typically require extensive datasets and high-quality labels to achieve optimal performance. While measurements are often readily available, the labeling process is frequently a major bottleneck. Active learning methods aim to address this by utilizing the fact that different data points contribute varying amounts of information to the model. However, these methods often make unrealistic assumptions, such as relying on an oracle to provide flawless labels at a consistent cost and effort. In this work, we propose new active learning strategies for classifying time series data, such as sensor measurements from fluctuating environmental conditions like electricity usage. These strategies are particularly useful for applications like home energy management, where labeling data can be a significant challenge. We introduce an innovative acquisition function that incorporates both model uncertainty and labeling uncertainty, while also addressing class imbalance. Additionally, we propose a stopping criterion that halts the active learning process once an optimal point is reached, minimizing the labeling effort. We also examine the impact of labeling errors on model performance and suggest two methods to reduce their effects: (i) a re-labeling approach based on label similarity, and (ii) a revised loss function that incorporates expert confidence levels. Our approach is validated through energy disaggregation tasks in a real-world setting with three domain experts. The results show that our methodology significantly enhances the performance of algorithms applied to new domains, reducing the number of labeled samples required—by up to 61% for dishwashers and 93% for kettles.*

***Keywords:** machine learning, datasets, labeling, active learning, time series, classification, uncertainty, energy management*

### **Introduction:**

The first part of the text describes the challenges and solutions in applying machine learning to time-series data, especially for electrical measurements in energy management applications. It explains that advanced machine learning algorithms often require large datasets with accurate labels to achieve optimal performance. However, while measurements are easy to collect, the labeling process is a bottleneck. Active learning methods address this by selecting the most informative data points for labeling, reducing the labeling effort. These methods exploit the fact that not all samples contribute equally to model performance. However, many active learning approaches make unrealistic assumptions, such as the availability of error-free labels at no extra cost.

The paper proposes a novel active learning approach for classifying time-series measurements, typically from sensors measuring fluctuating environmental conditions such as electricity consumption. This method is especially useful in home energy management, where labeling data can be difficult. The authors introduce an acquisition function that accounts for both model uncertainty and labeling uncertainty, ensuring balanced class distribution. A stopping criterion is also proposed, which halts the active learning process once the optimal point is reached, minimizing labeling effort. The paper also explores the impact of labeling errors and presents two methods to mitigate them: (i) a re-labeling mechanism based on label similarity, and (ii) a revised loss function that incorporates expert confidence levels.

The paper outlines the development of a human-in-the-loop active learning approach for time-series data, specifically focused on electrical measurements from non-intrusive load monitoring (NILM). The core idea is to address the challenges of labeling large datasets for machine learning models by integrating human feedback into the active learning loop. This method involves using human experts as a source of annotations, but unlike typical active learning models, it accounts for potential human errors in labeling. The process works by first training a model on a small labeled dataset, then iteratively selecting the most informative and diverse samples from a larger pool of unlabeled data. These selected samples are then labeled by experts and added to the model for further training.

The methodology includes a novel acquisition function, designed to balance class distribution while considering both model uncertainty and labeling uncertainty. To minimize unnecessary

labeling, a stopping criterion is introduced to halt the learning process once the model achieves satisfactory performance. The approach also addresses potential errors in labeling by implementing mechanisms for re-labeling and by incorporating expert confidence levels into the labeling process, ensuring that uncertain or noisy labels are appropriately handled.

In real-world applications, this approach was tested on energy disaggregation tasks using publicly available datasets. The experiments showed that the proposed method significantly reduces the amount of labeled data required while maintaining or even improving model performance, especially in scenarios where labeling errors could occur. This human-in-the-loop framework enhances the model's ability to generalize to new, unseen data, demonstrating its effectiveness in reducing labeling effort and improving model reliability in dynamic, real-world settings.

The paper introduces a human-in-the-loop active learning (HITL-ML) methodology applied to time-series data, specifically for non-intrusive load monitoring (NILM). This method is aimed at smart home energy management, addressing challenges related to labelling and model performance. The system involves a process where the machine selects data samples to query, and a human expert labels these samples. The expert also provides a confidence level for each label, which helps in handling potential errors in the labels. This approach ensures that the machine can learn incrementally until a stopping criterion is met.

The paper explores energy disaggregation, a process that separates a building's aggregate energy consumption into individual appliance usage. This fine-grained data can help with energy conservation and demand response strategies. While deep learning models are typically used in NILM, they require large labelled datasets for good performance, which are costly and time-consuming to obtain. The study acknowledges this challenge and introduces strategies for reducing the labelling effort.

By employing transfer learning techniques, the proposed method aims to improve the model's performance in new, unseen homes, even with minimal labelled data. Additionally, the paper discusses how to mitigate the impact of labelling errors through re-labelling and expert confidence mechanisms. These innovations reduce the need for extensive labelled data and enhance model stability and transferability across different environments.

This methodology is validated in real-world scenarios, using publicly available datasets and employing a deep neural network architecture designed for NILM. The results show that the approach significantly improves the performance of pre-trained models when applied to new homes, ensuring that even with limited labelled data, high-quality predictions can still be achieved.

In this study, the authors present an innovative active learning framework specifically designed for time-series electrical measurement data, typically gathered through non-intrusive load monitoring (NILM) systems used in energy disaggregation. The goal is to minimize the labeling effort required while maintaining high model performance. The process starts by training a model using a small set of labeled data, which is then expanded by selecting the most informative unlabeled samples for labeling through active learning. The acquisition function identifies which samples are most valuable to label, based on criteria such as uncertainty or informativeness.

The paper proposes a unique acquisition function that balances high-uncertainty samples with those that contribute to model diversity, ensuring that the model learns from a broad set of data. This acquisition function is designed to prioritize samples where the model's predictions are uncertain, but also includes samples that add variety to the dataset, particularly when the model starts to show high confidence. Additionally, a stopping criterion is introduced to prevent unnecessary labeling. The criterion terminates the active learning process once enough informative samples have been labeled, reducing the total labeling effort required.

Another key aspect of the approach is the use of expert confidence levels during labeling. By incorporating confidence scores from the experts, the model accounts for the potential variability in label quality, giving more weight to highly confident labels during training. The proposed methodology also includes a re-labeling mechanism that allows for correcting errors in labels based on the model's uncertainty, further enhancing the quality of the training data. The results demonstrate the effectiveness of these techniques in real-world NILM applications, where labeling data can be

both costly and time-consuming. The active learning strategy significantly improves model performance with fewer labeled samples, making it a valuable tool for applications in energy management and other domains requiring large datasets for model training.

The study evaluates the impact of labeling errors on the classification performance of time-series electrical data in the context of non-intrusive load monitoring (NILM). When errors are introduced during the labeling process, two main types of mistakes are considered: false positives and false negatives. The analysis reveals that false negatives—where an appliance activation is missed—have a more significant negative impact on the model's performance compared to false positives. This is because false negatives introduce additional negative samples that make it more difficult for the model to identify the correct patterns, while false positives do not drastically change the model's ability to recognize important appliance activations.

The effectiveness of a re-labeling mechanism to address these errors is demonstrated, with improvements in performance when mis-labeled samples are re-evaluated. The re-labeling mechanism targets samples that the model has identified as potentially mislabeled, based on a low match rate between predicted and actual labels. Once these samples are sent back for re-labeling, the accuracy of the dataset improves, which in turn enhances the model's classification performance.

Additionally, the study explores the use of expert confidence levels to reduce the impact of labeling errors. By incorporating confidence levels into the loss function during model training, the model adjusts its learning process based on the certainty of the expert's labels. This helps in mitigating the effects of potentially erroneous labels, especially when the data is noisy and the labels are less distinct. The results confirm that using confidence levels in the training phase improves model performance, particularly for appliances with less distinct activation patterns.

The experiments aim to evaluate the effectiveness of the proposed acquisition function, stopping criteria, and the human-in-the-loop active learning approach for classification of time-series electrical measurement data, particularly in the context of non-intrusive load monitoring (NILM). The primary goal is to assess the impact of the new acquisition function against traditional benchmarks and to test the re-labelling mechanism's effectiveness.

In the first experiment, the focus is on transfer learning with labels obtained via submetering, where samples from a query pool are labeled using submetered electricity consumption measurements. This experiment simulates labeling errors, including false positives and false negatives, to evaluate how the proposed approach handles inaccuracies. By applying the acquisition function, which incorporates model uncertainty and class balancing, the active learning process selects the most informative samples for labeling. This ensures the model learns from diverse data points, and the re-labelling mechanism can correct any erroneous labels. Additionally, simulated confidence levels are used to mitigate the effects of labeling errors, emphasizing the importance of expert confidence when providing labels.

The second experiment moves from simulated to real-world conditions, where experts provide labels during the active learning process. In this case, the quality of the labels is assessed based on the expert's confidence level, assuming that lower confidence correlates with a higher likelihood of error. This setup mimics a real-world scenario where labeling can be subjective and error-prone. The results highlight the importance of incorporating expert confidence into the learning process, demonstrating that models can be significantly improved even with a relatively small number of labeled samples.

Both experiments utilize a user interface developed to assist experts in the labeling process. This interface displays time-series data, allowing experts to quickly annotate windows of electrical consumption with confidence levels. The user-friendly design helps facilitate accurate and efficient labeling, which is crucial for the success of the active learning approach in real-world applications.

The results from both experiments show that the proposed methods effectively reduce labeling effort, improve model performance, and ensure the model can adapt to new, unseen data with minimal labeling.

The study explores the performance of different acquisition functions used in active learning for non-intrusive load monitoring (NILM). Specifically, it compares the proposed acquisition

function, which incorporates model uncertainty and class balancing, with well-established benchmarks such as pool-based and stream-based uncertainty, BADGE, and CLUE. The results show that the proposed method outperforms other functions in terms of F1-score, especially for appliances with challenging signatures, such as kettles and microwaves, in the REFIT house 5 dataset.

The experiments also highlight the effectiveness of the proposed stopping criterion, which minimizes labeling effort while maintaining performance. By halting the active learning process once the model reaches an optimal performance point, this criterion significantly reduces the need for further labeling with only a slight performance trade-off. This early stopping method is shown to save considerable effort in labeling without compromising the model's accuracy.

Additionally, the paper examines the impact of labeling errors, particularly false negatives and false positives, on model performance. It is demonstrated that false negative errors (mislabeling appliance activations as non-activations) are particularly detrimental, especially for appliances with short activation durations, such as kettles. On the other hand, false positive errors (mislabeling non-activations as activations) are less impactful, as they typically do not cause the model to forget previously learned patterns. The study further evaluates a re-labeling mechanism that improves model performance by correcting erroneous labels, particularly when false negatives are present.

Finally, the integration of expert confidence into the labeling process is tested. This mechanism assigns greater weight to labels given by experts with higher confidence, thus improving the reliability of the labeled data and enhancing the model's performance. The results from these experiments provide valuable insights into how human-in-the-loop strategies can significantly improve active learning for NILM, even in the presence of labeling errors.

The experiments evaluate the performance of the proposed active learning acquisition function by comparing it to established benchmarks. In this study, different methods such as pool-based uncertainty, stream-based uncertainty, BADGE, and CLUE are tested. The proposed method, which incorporates both model uncertainty and class balancing, shows comparable results to these benchmarks, achieving high F1-scores while reducing the number of labeled samples. Notably, the proposed method efficiently handles the challenge of selecting informative samples, balancing the trade-off between early-stage performance improvement and long-term learning stability.

Additionally, the paper introduces a stopping criterion designed to minimize labeling effort once the model has achieved optimal performance. This approach effectively reduces the need for continued labeling while maintaining accuracy, showing minimal performance loss after halting. The stopping points are determined based on the number of high-uncertainty samples present in the query pool. This method significantly saves labeling resources without compromising the performance of the model.

Furthermore, the impact of labeling errors is explored, specifically focusing on false negative and false positive errors. The study shows that false negatives—incorrectly labeling appliance activations as non-activations—have a larger negative impact on performance, particularly for appliances with brief activation periods. The re-labeling mechanism introduced in the study helps correct these errors, thus enhancing the model's robustness and accuracy. Lastly, the inclusion of expert confidence levels in the labeling process further improves model performance, especially in the presence of noisy or challenging samples. This approach is validated through real-world experiments, where even with labeling errors, the model shows improved performance through the use of active learning strategies.

The experiment results, shown in figures and tables, compare the performance of the proposed active learning acquisition function with various state-of-the-art methods. The main goal is to reduce labeling efforts while maintaining or improving model accuracy for non-intrusive load monitoring (NILM). The results reveal that the proposed function significantly reduces labeling effort by selecting more informative samples and stopping the process at an optimal point. Specifically, labeling effort is reduced by up to 93% for certain appliances in the UK-DALE dataset, and between 61% and 88% in the REFIT dataset.

The results also demonstrate that even when errors, such as false positives and false negatives, are introduced into the labeling process, the active learning methodology continues to enhance model



generalization across different datasets. The proposed re-labeling mechanism further improves the model by identifying potentially erroneous labels and correcting them. Moreover, incorporating expert confidence into the process prevents a decline in performance due to the accumulation of inaccurate labels. These results validate the effectiveness of the human-in-the-loop approach, particularly in real-world scenarios where labeling can be error-prone.

Furthermore, the performance of the acquisition function is compared to traditional methods like pool-based and stream-based uncertainty functions, as well as more advanced techniques such as BADGE and CLUE. The proposed approach shows competitive or superior results, especially in terms of achieving high F1-scores with fewer labeled samples. The stopping criterion also plays a crucial role in preventing unnecessary labeling, showing that early stopping leads to minimal performance loss with significant resource savings.

**Tomaž Brzin, Miha Brojan** (*University of Ljubljana, Ljubljana, Slovenia*)

## **GAN-BASED INVERSE DESIGN OF SOFT MORPHING COMPOSITE BEAMS.**

**Abstract:** *Designing structures with customized properties presents a significant challenge due to the multitude of potential solutions that can meet specified criteria. In the case of morphing composite beams, for instance, various fabrication strategies are possible, as parameters such as material, geometry, and actuation mechanisms can be altered. Additionally, these problems often involve strong nonlinearities arising from large deformations. To address these complexities, we introduce a generative adversarial network (GAN)-based inverse design framework tailored for the development of soft composite beams capable of morphing into predefined shapes and executing intricate motions. Our method leverages composite materials comprising both passive and active layers, which deform into target configurations due to strain mismatches driven by non-uniform geometric and material properties combined with temperature-induced actuation. To expedite the exploration of the design parameter space, we devised a mechanical analog — a simplified "toy model" — that replaces thermal actuation with the mechanical stretching of highly elastic active layers. This analog allows for faster testing and evaluation of design parameters. Through experimental and numerical validation, we demonstrated the efficacy of our approach. The generator component of the GAN takes target shapes as inputs and outputs the corresponding fabrication parameters for creating composite beams that autonomously deploy into desired configurations upon release. Furthermore, we extended this methodology to produce design parameters for composite beams capable of complex, temperature-actuated movements. Our data-driven framework is both straightforward and robust, offering solutions to intricate design challenges. It holds promise for advancing the development of soft robotics and smart deployable structures, paving the way for innovative applications in these domains.*

**Keywords:** *inverse design, composite beams, generative adversarial networks, morphing structures, strain mismatch, active layers, soft robotics, deployable structures*

### **Introduction:**

The design of structures with specific, customized properties often presents significant challenges, primarily due to the existence of multiple possible solutions that can fulfill the same set of prescribed conditions. This complexity becomes even more pronounced in scenarios such as the creation of morphing composite beams. In these cases, the choice of materials, geometric configurations, and actuation methods introduces a broad spectrum of potential fabrication solutions. Compounding this challenge is the inherent nonlinearity of such problems, driven by the large deformations that occur during the structural transformation.

To tackle these issues, we propose a novel method based on generative adversarial networks (GANs) to facilitate the inverse design of soft composite beams. These beams are engineered to morph into predefined shapes and perform intricate, prescribed motions. Our approach incorporates the use of layered composite materials, consisting of both passive and active layers. The deformation into target shapes is achieved through strain mismatches caused by non-uniform geometric and material properties, in conjunction with temperature-induced actuation.

To expedite the exploration of the parametric space and streamline the validation of our method, we developed a simplified mechanical analog. This "toy model" uses mechanical stretching of highly elastic active layers to mimic the effects of heating and cooling. Through experiments and simulations, we demonstrate the efficacy of this model. The generator component of the GAN receives target shapes as inputs and outputs corresponding design parameters, enabling the fabrication of composite beams that self-deploy into the desired configurations upon release.

The second part of the text describes the concept and challenges of designing deployable beams with predefined shapes. In this approach, the target final configurations of the structure are known in advance, while the fabrication parameters required to achieve these configurations remain to be determined.

The foundational idea involves starting with a flat, thin elastic material layer. This base layer is then bonded with pre-stretched layers, which are strategically placed either on its top or bottom surface to induce specific curvatures. These pre-stretched layers store elastic potential energy that drives the deformation of the composite beam into its intended shape once released. The design

parameters—such as the amount of pre-stretch, layer dimensions, and material properties—must be carefully optimized to achieve the desired outcomes.

A key innovation of this method is training a neural network model, specifically a generator network, to predict the required design parameters for the target shapes. These parameters include the amount of pre-stretch, the dimensions of each layer, and the material's modulus of elasticity. The generator network accepts the target shape as input and provides a detailed set of fabrication parameters that ensure the structure morphs into the intended configuration.

Although this model currently focuses on single-use deformations (non-reversible morphing), it provides a practical proof-of-concept for designing deployable structures. The same principles could be extended to reversible mechanisms, such as heat-actuated systems, where the structure returns to its original state upon cooling.

This simplified system not only allows for the rapid testing of design principles but also lays the groundwork for more complex applications in smart materials and deployable systems, including soft robotics and adaptive structures.

Next part introduces the generative adversarial network (GAN)-based model developed for the inverse design of morphing structures. The GAN architecture comprises three interconnected neural networks: the generator, the critic, and a pre-trained simulator. These components work together to produce feasible design parameters that enable the fabrication of deployable composite beams.

The generator network is responsible for creating candidate solutions for the design parameters based on the input target shapes. It takes a normalized representation of the target curvature, combined with random noise, to explore diverse design possibilities. The critic network evaluates the feasibility of these generated parameters by comparing them to a library of known feasible design samples. Through this adversarial process, the generator learns to produce realistic and functional outputs.

To enhance accuracy, the GAN framework integrates a pre-trained simulator network. This simulator predicts the expected shape of the composite beam based on the proposed design parameters. By incorporating this additional layer, the model ensures that the generated solutions closely align with the target shapes. The overall training process balances two key objectives: feasibility (ensured by the critic network) and shape accuracy (guided by the simulator network).

The generator network architecture includes dense layers interspersed with batch normalization and activation functions, designed to output five critical design parameters. These parameters include layer dimensions, pre-stretch values, and material elasticity. The critic network, with its leaky ReLU activations, focuses on identifying non-feasible designs and guiding the generator to refine its outputs. The simulator network, resembling an inverted generator structure, bridges the gap between the design parameters and the physical realization of the target shapes.

The training dataset, created specifically for this model, includes thousands of pairs of design parameters and corresponding shape representations. This comprehensive dataset enables the neural networks to learn the intricate relationships between input design variables and the resulting beam shapes. Training involves alternating updates to the generator and critic networks, ensuring convergence to a reliable solution.

This innovative approach demonstrates the potential of GAN-based frameworks to address complex inverse design challenges. By generating accurate and feasible design parameters, the model enables the rapid prototyping and fabrication of morphing composite beams with applications in soft robotics, adaptive systems, and other advanced engineering domains.

This part focuses on the experimental fabrication of morphing composite beams using the design parameters generated by the GAN-based model. The fabrication process begins with creating silicone sheets of varying thicknesses using 3D-printed molds. These sheets serve as the foundational materials for constructing composite beams.

Each active layer is carefully cut and stretched according to the generated design parameters, which dictate the required dimensions and pre-stretch values. The stretched layers are then bonded to a base layer using a silicone-based polymer adhesive to ensure a strong and durable connection. To

maintain alignment during bonding, an apparatus consisting of 3D-printed parts, threaded rods, and nuts is employed. This setup allows precise control over the stretching and placement of the active layers.

The fabrication process involves positioning the pre-stretched layers on the base layer, applying adhesive to the segment length, and securing the assembly with an acrylic plate and weights to prevent misalignment. Once the adhesive cures, excess material is trimmed, and the composite beam is released to deform into its target shape.

This methodical approach ensures that the fabricated beams closely resemble the target configurations generated by the GAN model. The results validate the effectiveness of the model in producing design parameters that translate into physical structures with the intended morphing capabilities. Moreover, this fabrication workflow serves as a scalable and efficient process for developing soft, deployable composite beams for advanced applications.

The process began by defining a set of target shapes, including complex configurations such as waves, spirals, and circular forms, which served as inputs for the trained GAN model. The generator network produced the design parameters necessary to fabricate these shapes, demonstrating its ability to handle diverse and intricate geometries.

Fabricated composite beams were optically scanned and analyzed to compare their final configurations with the target shapes. The results revealed excellent alignment between the predicted and fabricated shapes, with minimal deviations. These deviations were attributed to minor inaccuracies during the fabrication process, such as slight errors in cutting, bonding, or stretching the active layers. Despite these challenges, the generated parameters consistently led to high-quality morphing beams.

To further validate the model, creative shapes inspired by line art and handwritten designs were used. The GAN model successfully generated parameters for these unique configurations, showcasing its versatility. Additionally, the efficiency of the approach was highlighted by its ability to generate thousands of design parameter combinations in seconds, a significant improvement over traditional methods like evolutionary algorithms.

The study also included numerical comparisons of curvature radii, showing strong agreement between the generated, simulated, and experimentally measured values. This robust performance underscores the model's capability to handle complex inverse design problems, ensuring that the fabricated beams meet precise specifications.

Overall, the results demonstrate the potential of this GAN-based approach for designing and fabricating deployable composite structures with practical applications in areas such as soft robotics, adaptive systems, and advanced engineering solutions.

Next part explores the extension of the GAN-based framework to design morphing composite beams capable of achieving complex, temperature-actuated motions. This advanced implementation replaces the pre-stretch parameter with the coefficient of linear thermal expansion, allowing the beams to morph in response to controlled heating or cooling.

In this scenario, active layers are positioned either above or below the base layer, with their deformation controlled by thermal loads. For instance, cooling the active layer on top induces positive curvature, while heating it produces the opposite effect. This adaptation enables the creation of beams that transition seamlessly between two or more predefined shapes based on temperature changes.

To demonstrate this capability, two morphing beams were designed. The first beam shifted from a home position at a lower temperature to an end position at a higher temperature. Similarly, the second beam showcased continuous motion between two target shapes as the temperature was gradually increased. The generated design parameters for these configurations ensured consistent geometric and material properties across all segments, with temperature being the sole variable.

Numerical simulations and experimental validation confirmed the model's effectiveness. Despite minor deviations caused by inherent fabrication constraints, the results closely matched the target configurations. This demonstrates the model's ability to generate reliable and accurate design parameters for complex morphing motions.

This extended approach highlights the versatility and robustness of the GAN-based framework in addressing diverse inverse design challenges. It paves the way for innovative applications in soft robotics, smart materials, and deployable systems, where temperature-actuated motions play a critical role in functionality and adaptability.

Provides a summary of the research and highlights the achievements of the proposed method for designing morphing composite beams. The study addressed the inverse design problem, where the desired target shape of a self-deployable composite beam was predefined, but the necessary fabrication parameters were unknown. By employing a machine-learning-driven approach, the research introduced a robust generative adversarial network (GAN)-based framework to tackle this challenge.

The core of the method involved three interconnected neural networks: a generator to propose design parameters, a critic to ensure feasibility, and a simulator to validate the accuracy of these parameters. Together, these networks worked seamlessly to generate fabrication guidelines for composite beams that morph into specified shapes upon deployment. This data-driven model leveraged strain mismatch in composite materials as a source of potential energy, eliminating the need for external stimuli to trigger the shape transformation.

The experimental validation showcased the effectiveness of this approach. Several composite beams were fabricated using the generated parameters, and their shapes closely aligned with the predefined targets. The minimal deviations observed were primarily attributed to practical constraints during the fabrication process, such as minor inaccuracies in layer bonding or material handling.

The study also explored more complex applications, including temperature-actuated morphing beams capable of transitioning between multiple shapes. This extension demonstrated the adaptability of the GAN-based framework in addressing diverse and challenging design problems.

The proposed method opens new avenues for creating advanced morphing structures, including deployable systems and soft robotics, that combine simplicity in fabrication with high functionality. Its ability to handle complex inverse design tasks with speed and precision underscores its potential for practical implementation in engineering and materials science.

*Наукове видання*

За загальною редакцією  
д-ра техн. наук, професора

**КОВАЛЕВСЬКОГО Сергія Вадимовича**

**НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ  
ТА ЇХ ЗАСТОСУВАННЯ  
НМТІЗ-2024**

**Збірник наукових праць**

За авторським редагуванням

32/2016    Формат 60 x 84/16.    Ум. друк. арк. 20,2  
Обл.-вид. арк. 20,6. Тираж 100 прим. Зам. № 25

Видавець і виготівник  
Донбаська державна машинобудівна академія  
84313, м. Краматорськ, вул. Академічна, 72.  
Свідоцтво суб'єкта видавничої справи  
ДК № 1633 від 24.12.2003